

Nanocarbons for Supercapacitors

FRANÇOIS BÉGUIN, ENCARNACIÓN RAYMUNDO-PIÑERO
Research Centre on Divided Matter,
CRMD-CNRS & Orléans University,
Orléans, France

Article Outline

Glossary
Definition of the Subject
Introduction
Brief Overview of Activated Carbons Properties
General Properties of Electrochemical Capacitors
Carbons for Electrical Double-Layer Capacitors
Carbons for Pseudo-capacitors
Conclusion and Future Directions
Bibliography

Glossary

Carbon structure and texture The structure is the arrangement of the carbon atoms in the space since the texture is the arrangement of the graphene layers in the space for giving porosity or empty space accessible for molecules or ions.

Electric double layer capacitor Electrochemical capacitor in which the charge storage is achieved electrostatically because of the separation of charges in a double layer across the electrode/electrolyte interface.

Nanocarbon Carbon material consisting in more or less disordered graphene layers which can be synthesized with different structures, porous texture, and surface functionality.

Pseudo-capacitor Electrochemical capacitor in which the charge storage is achieved by an electron transfer that produces chemical or oxidation state changes in the electrode materials. As a difference from a battery, the electrode potential varies proportionally to the charge exchanged.

Supercapacitor or electrochemical capacitor Electrochemical energy storage device in which the voltage declines linearly with the extent of charge. A supercapacitor consists of two electrodes separated by a porous membrane immersed in an electrolyte.

Surface functionality Surface groups are ubiquitously present at the edges of graphene sheets in carbon materials, especially the high surface area ones. Oxygen-containing surface groups are the most occurring ones but carbon can contain other heteroatoms, such as nitrogen or sulfur.

Definition of the Subject

Supercapacitors (or electrochemical capacitors) are electrochemical energy storage devices having higher energy density than dielectric capacitors and higher power density than batteries. Actually, they are capable of delivering large amounts of energy in a very short time. These devices rely mainly on the characteristics of the electrical double layer that forms on all polarized conductors when immersed in an electrolyte. The double layer forms in less than 10^{-6} s and responds to changes in a similar time frame. This is about 1,000 times faster than an electrochemical reaction at a battery electrode that has a time constant in the range of 10^{-3} s. These devices find application where high-power delivery is required.

Introduction

Supercapacitors (or electrochemical capacitors or ultracapacitors) represent a relatively new energy storage technology which has known an important development during the recent years. Supercapacitors can be applied for stationary and mobile systems where high-power pulses are requested: cars acceleration, tramways, cranes, lifts, safety systems, etc. Moreover, owing to their low time constant, they can quickly harvest energy, for example during deceleration or

braking of vehicles. Although supercapacitors are able to provide high power with a long cycle life compared to accumulators, they suffer of a relatively low energy density. Therefore, the main ongoing research direction concerns the optimization of the existing electrode materials and the development of new materials, both for protic electrolytes and aprotic organic electrolytes.

Industrial supercapacitors are essentially based on nanoporous carbon electrodes. The reasons of the choice lie in the high availability, low cost, chemical inertness, and good electrical conductivity of activated carbons, as well as a high versatility of texture and surface functionality. For these reasons, this chapter will present the capacitance properties of carbon-based electrodes showing optimization strategies playing on the structure/nanotexture of carbon and the nature of the electrolyte.

Brief Overview of Activated Carbons Properties

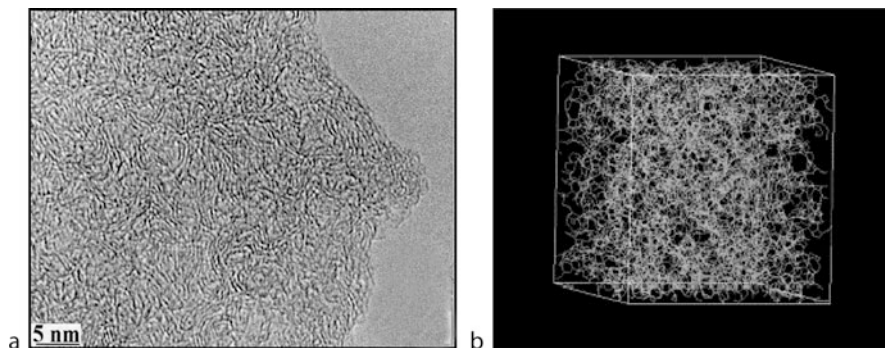
In industry, activated carbons are essentially produced by carbonization (pyrolysis at temperatures up to 900°C under neutral atmosphere) of various precursors (lignocellulosic, polymers, anthracites, ...), followed by physical and chemical activation [1]. Physical activation is generally realized around 900°C through partial gasification of carbon, using CO₂ or steam, according to Eqs. 1 and 2:



For chemical activation, the main reagents used are KOH, ZnCl₂, and H₃PO₄ [1]. KOH activation is a complex process, involving several redox reactions with carbon, followed by potassium intercalation/insertion and expansion of the structure [2, 3]. In the particular case of H₃PO₄, carbonization and activation generally proceed simultaneously at temperatures lower than 600°C [4].

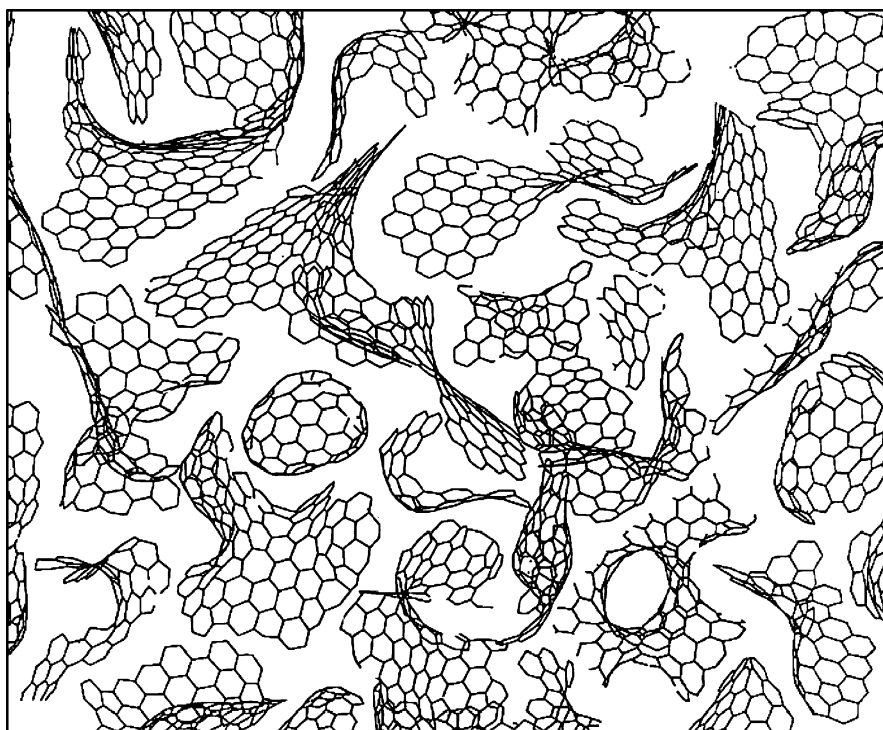
On the nanotextural point of view, activated carbons (AC) are characterized by high values of specific surface area, up to 3,000 m² g⁻¹. Depending on the precursor and carbonization/activation conditions, they can be essentially microporous (pores < 2 nm) or microporous and mesoporous (2 nm < pores < 50 nm). Many structural/nanotextural models have been proposed, but none of them gives a perfect description of the properties. The HRTEM 002 lattice-fringe image of an AC in Fig. 1a shows short fringes, few nanometers long, representing graphene-type layers, highly misoriented, some of them forming stacks of few layers. The disorientation of the graphene-type units causes the porosity of the material. According to simulations (Fig. 1b, [5, 6]), the graphene layers are not perfectly planar, and curvatures might be produced by defects such as pentagons and heptagons in their structure.

Based on various observations, Harris et al. proposed a model for the structure/nanotexture of non-graphitizable carbons that consists of discrete fragments of curved carbon sheets, in which pentagons and heptagons are dispersed randomly throughout networks of hexagons, as illustrated in Fig. 2 [7–9]. The size of the micropores in this model would be of



Nanocarbons for Supercapacitors. Figure 1

(a) HRTEM 002 lattice fringe image of an activated carbon; (b) structure/nanotexture of an activated carbon obtained by coupling HRTEM and numerical simulation [6]



Nanocarbons for Supercapacitors. Figure 2

Schematic nanotextural and structural model of an AC showing fullerene-like elements [7–9]

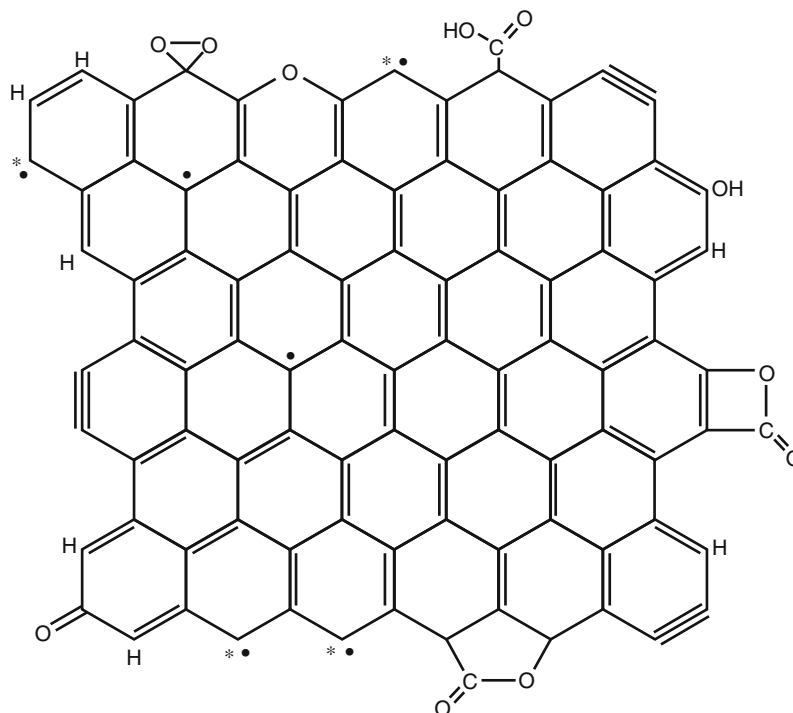
the order of 0.5–1.0 nm, which is similar to the pore size observed in typical microporous activated carbons. Taking into account that the independent fullerene-like graphene units would be soluble in some solvents and that the thermal stability of activated carbons is high, one has to imagine that, in addition to this representation, the units are connected by bridges which keep the backbone.

Apart of carbon which is the main element constituting nanoporous carbons, heteroatoms (essentially oxygen and nitrogen) in the amount of few atomic percent, generally less than 5 at.%, are always present in activated carbons. These heteroatoms are either remains of the carbon precursor and/or they are introduced during the activation process. Figure 3 shows a representation of the main oxygenated functionalities found at the edge of graphene planes [10]. Beside, free edge sites are present together with unpaired electrons stabilized by resonance. In the case of C/C supercapacitors in organic electrolyte, the later are assumed to be responsible of electrolyte decomposition, leading to capacitance decrease and resistance

increase of the systems during cycling [11]. Therefore, specific treatments of carbons are mandatory to partly deplete these sites and get rid of the drawbacks linked to their existence.

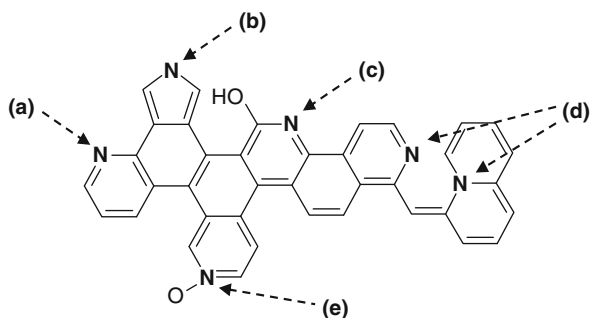
In the case of nitrogen, the variety of functionalities (Fig. 4) results both from the position occupied in the ring system and from the extent of association with oxygen, which hardly can be avoided during synthesis. Nitrogen is either substituted to carbon (“lattice nitrogen”) or included in the form of functional groups (“chemical nitrogen”) at the periphery of polyaromatic structural units [12, 13]. The surroundings of the nitrogen atom in a graphene layer obviously affect its charge, electron donor/acceptor properties, and the contribution to the delocalized π electron system.

For supercapacitor carbon electrodes, it will be further shown that (1) the developed surface area is responsible of an important electrical double-layer capacitance; (2) both the oxygenated and nitrogenated functionalities may be involved in redox reactions with the electrolyte, which enhance capacitance through a pseudo-capacitive contribution.



Nanocarbons for Supercapacitors. Figure 3

Schematic representation of some oxygenated functionalities present at the edges of a graphene unit. Free edge sites are present in addition to surface functionalities. In ambient, unpaired electrons are stabilized by resonance on the basal plane [10]



Nanocarbons for Supercapacitors. Figure 4

Nitrogen functionalities in a carbon material. (a) Pyridinic (N-6); (b) pyrrolic; (c) pyridonic (N-5); (d) quaternary (N-Q); and (e) oxidized nitrogen (N-X)

General Properties of Electrochemical Capacitors

An electrochemical capacitor is generally constituted of two identical carbon electrodes (carbon coating a current collector), separated by a porous membrane,

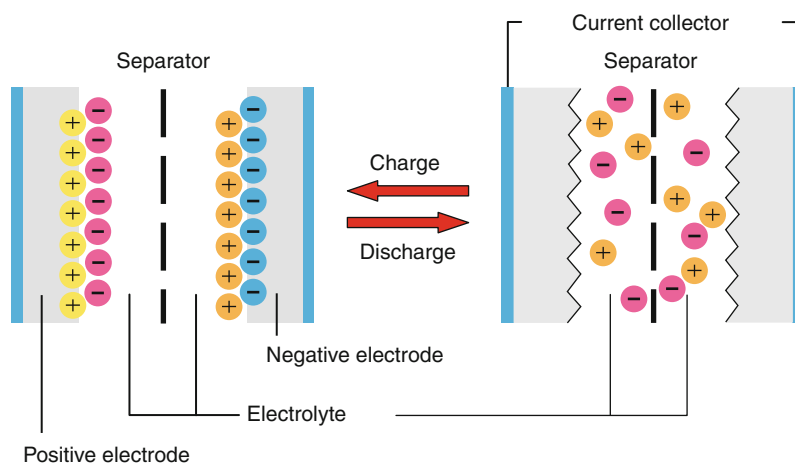
and immersed in an electrolyte (Fig. 5). When the system is charged, the anions of the electrolyte are electrosorbed on the positively polarized electrode and the cations on the negative one, giving rise to an electric double layer at each electrode. As shown in Fig. 5, in its charged state, this system is equivalent to two capacitors of capacitance C_1 and C_2 in series, and the resulting capacitance C is given by the classical relationship (Fig. 3):

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} \quad (3)$$

The capacitance originating itself from the electrical double layer at each electrode/electrolyte interface is given by the formula (Eq. 4) [14]:

$$C_e = \epsilon A / d \quad (4)$$

where ϵ is the dielectric constant or permittivity, A the surface area of the electrode/electrolyte interface, and d the distance between the two charged layers of



Nanocarbons for Supercapacitors. Figure 5

Discharged (*right*) and charged (*left*) states of an electrochemical capacitor

opposite sign. The value of d is controlled by the size of ions and is of the order of 1 nm. Consequently, the capacitance originating of the electrical double layer is very high, e.g., around 0.1 F m^{-2} . In electrical double-layer capacitors (EDLC), the electrodes are from high surface area-activated carbons; for this reason, the surface area of the interface is very high and the capacitance C_e reaches values as high as 100 F per gram of carbon.

As said in the introduction, the main research direction on supercapacitors is dedicated to enhancing the energy density of these systems in order to broaden their possible applications. The energy density of supercapacitors is given by formula (Eq. 5):

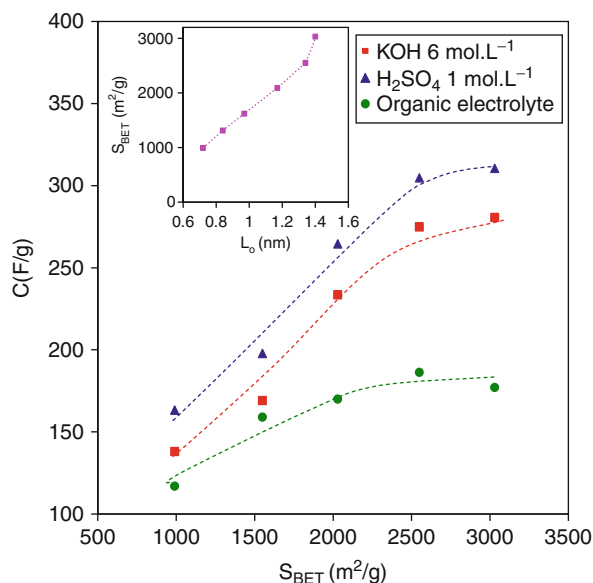
$$E = \frac{1}{2} CU^2 \quad (5)$$

where U represents the operating voltage. In supercapacitors, the later is strongly depending on the stability window of the electrolyte. In water medium, although the thermodynamic value is 1.23 V, practically it ranges from 0.7 to 1 V [15]. In organic electrolyte, the operating voltage can reach 2.7–2.8 V [15], and due to the square dependence of energy with voltage, this medium is definitively preferred for the industrial systems. Hence, the energy density of electrochemical capacitors is mainly controlled by the capacitance of the electrode materials. In aprotic electrolyte (e.g., organic electrolyte), the capacitance originates exclusively from the electrical double-layer formation,

whereas an additional pseudo-capacitive contribution, due to fast pseudo-faradic reactions between the electrode surface and the electrolyte, may appear in protic electrolytic media. These two contributions will be considered separately.

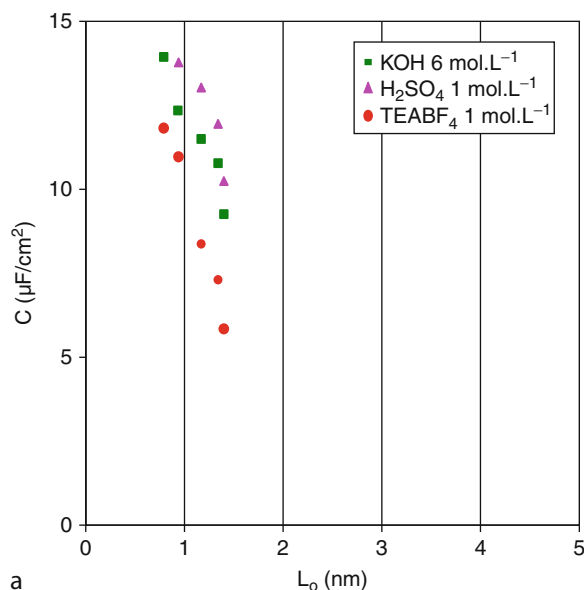
Carbons for Electrical Double-Layer Capacitors

Considering Eq. 4 suggests that developing the surface area of activated carbons should lead to an increase of capacitance. For this reason, in the literature, many authors have suggested a proportional dependence between the gravimetric capacitance and the BET specific surface area measured by nitrogen adsorption at 77 K. An example is shown in Fig. 6 for a series of nanoporous carbons obtained by KOH activation of bituminous coal pyrolyzed at various temperatures [16]. Although these carbons were prepared from the same precursor using the same activation process and should be considered of identical nanotexture, the proportionality is observed only for the low values of BET surface area. Above $2,000 \text{ m}^2 \text{ g}^{-1}$, the observed saturation phenomenon is often interpreted by the fact that the BET model overestimates the values of surface area [17, 18]. Therefore, some plots have been proposed versus the DFT (density functional theory) surface area, but saturation is still observed for the high surface area carbons, i.e., for carbons with increasing activation degree. In order to explain the saturation, Barbieri et al. suggested that due to the decrease of average pore



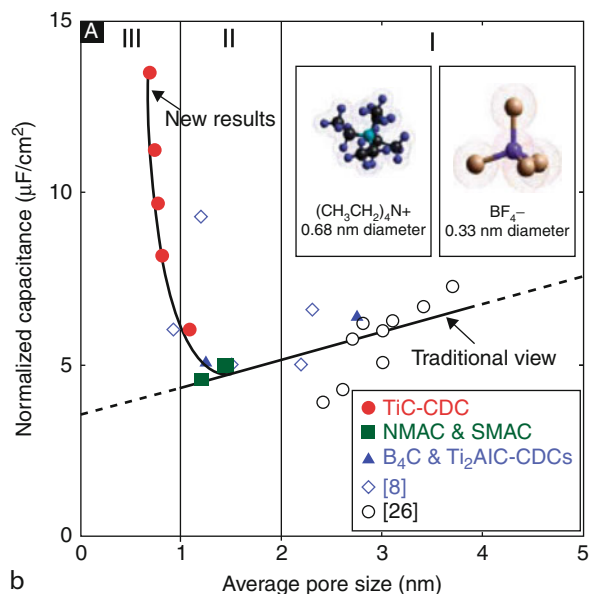
Nanocarbons for Supercapacitors. Figure 6

Gravimetric capacitance in various electrolytes (1 mol L⁻¹ H₂SO₄, 6 mol L⁻¹ KOH and 1 mol L⁻¹ TEABF₄ in acetonitrile) vs BET specific surface area for a series of bituminous coal-derived carbons activated by KOH at 800°C. The inset shows the relationship between S_{BET} and the average pore size L_0 (Adapted from [16])



wall thickness in highly activated carbons, the electric field (and the corresponding charge density) no longer decays to zero within the pore walls [19].

Another explanation can be proposed from inset of Fig. 6, showing that the average pore size increases together with the specific surface area, i.e., when the activation degree of carbon increases. In high surface area carbons, with larger pores, one might suggest that the interaction of ions with pore walls is weaker, leading to smaller values of capacitance. Therefore, the normalized capacitance, e.g., the capacitance divided by the DFT surface area, has been plotted versus the average size L_0 , both for the series of carbons from bituminous coal (Fig. 7a) [16] and for carbide-derived carbons (Fig. 7b) [20]. In the two cases, an increase of normalized capacitance is observed when the average pore size decreases, the highest values being observed for L_0 smaller than 1 nm. Taking into account the dimensions of solvated and non-solvated ions for tetraethylammonium tetrafluoroborate in acetonitrile, e.g., 1.30 and 0.67 nm for TEA⁺, 1.16 and 0.48 nm for BF₄⁻, respectively, suggests that the ions are at least partly desolvated in the pores [21]. Such property was already suggested for templated carbons which showed



Nanocarbons for Supercapacitors. Figure 7

Normalized capacitance versus average pore size (a) for bituminous coal-derived activated carbons in 1 mol L⁻¹ H₂SO₄, 6 mol L⁻¹ KOH, and 1 mol L⁻¹ TEABF₄ in acetonitrile (Adapted from [16]) and (b) for carbide-derived carbons (CDC) in 1 mol L⁻¹ TEABF₄ in acetonitrile [20]

proportionality between capacitance and the ultramicropore (<0.7–0.8 nm) volume [22].

At this step, it is now clear that adapting the pore size of carbons to the size of ions is more important than developing the pore volume. Indeed, high surface area carbons display generally broader pores, and according to the data presented above the normalized capacitance is not optimized. Hence, the trend should be to develop carbons with pores ranging between 0.7 and 1 nm, which are able to accommodate desolvated ions, while allowing them to optimally interact with the pore walls. In addition, with such carbons, the useless volume would be reduced, and the volumetric capacitance optimized, fulfilling the industrial requirements for getting compact systems.

However, it has been shown that using carbons with pores in the optimal range, from 0.7 and 1 nm, may lead to porosity saturation before reaching the maximum voltage for the considered electrolyte [23]. Indeed, the comparison of two carbons, named PC ($S_{\text{DFT}}=1,434 \text{ m}^2 \text{ g}^{-1}$) and VC ($S_{\text{DFT}}=2,160 \text{ m}^2 \text{ g}^{-1}$), with different pore size distribution (Fig. 8), in 1.5 mol. L^{-1} tetraethylammonium tetrafluoroborate in acetonitrile, shows a perfectly rectangular voltammogram in the whole voltage range for VC while the capacitive current dramatically decreases above around 1.5 V for PC (Fig. 9). According to Fig. 8, all pores of PC are narrower than 1 nm and consequently able to optimally interact with ions. However, if one considers the pores larger than the diameter of desolvated TEA^+ cations

(0.68 nm), the corresponding DFT values are 198 and $964 \text{ m}^2 \text{ g}^{-1}$, respectively, for PC and VC. From these values, the maximum theoretical charge able to be accommodated in the pores has been calculated for both carbons and compared to the charge determined by integration of the respective voltammograms [23]. For PC the theoretical and experimental values are almost identical, confirming that the shape of the voltammetry curve is due to porosity saturation. By contrast, in the case of VC, the theoretical value is larger than the experimental one, demonstrating that for this carbon the porosity is not saturated, at least for the maximum voltage value reached in this experiment.

In summary, the porous texture of carbons strongly influences their electrical double-layer properties. With TEABF_4 in acetonitrile, the normalized capacitance is optimal in subnanometric pores, which also indicates that ions are partly desolvated in charged electrodes. For obtaining high values of volumetric capacitance, nanoporous carbons should present high values of specific surface area ($\sim 2,000 \text{ m}^2 \text{ g}^{-1}$) with a narrow pore size distribution between 0.7 and 1 nm. Therefore, there is a need of new production techniques enabling to better reach this objective. Beside, taking into account that electrochemical intercalation is possible during charging [24], the structural parameters of carbons should also be more carefully considered in the future.

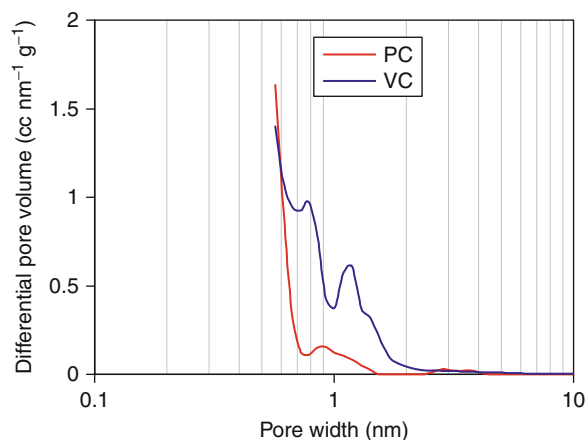
Carbons for Pseudo-capacitors

In general, two modes of energy storage are combined in electrochemical capacitors: (1) the electrostatic attraction between the surface charges and the ions of opposite charge (electrical double layer); (2) a pseudo-capacitive contribution which is related with quick redox reactions between the electrolyte and the electrode [14, 15]. Whereas the redox process occurs at almost constant potential in an accumulator, the electrode potential varies proportionally to the charge-exchanged dq in a pseudo-capacitor, what can be summarized by formula (Eq. 6):

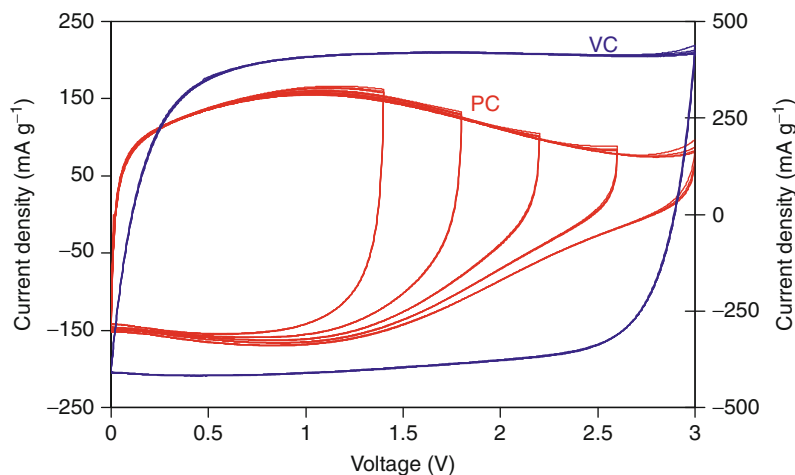
$$dq = C * dU \quad (6)$$

where dU represents the potential variation.

The electrical response of such a system is comparable to that of a capacitor. Being of faradic origin and



Nanocarbons for Supercapacitors. Figure 8
Pore size distribution of carbons PC and VC [23]



Nanocarbons for Supercapacitors. Figure 9

Cyclic voltammograms for the EDLCs based on PC carbon (*left-hand side* Y-axis for current) and VC carbon (*right-hand side* Y-axis for current) [23]

non-electrostatic, this capacitance is distinguished from the double-layer one and is called pseudo-capacitance. In summary, the electrical double-layer formation is a universal property of a polarized material surface, and pseudo-capacitance is an additional property which depends both on the type of electrode material and electrolyte. Compared to the double-layer normalized capacitance ($\sim 10 \mu\text{F cm}^{-2}$), it has generally a high value ($100\text{--}400 \mu\text{F cm}^{-2}$), because it involves the bulk of the electrode and not only the surface. From a practical point of view, pseudo-capacitance contributes to enhancing the capacitance of materials and their energy density.

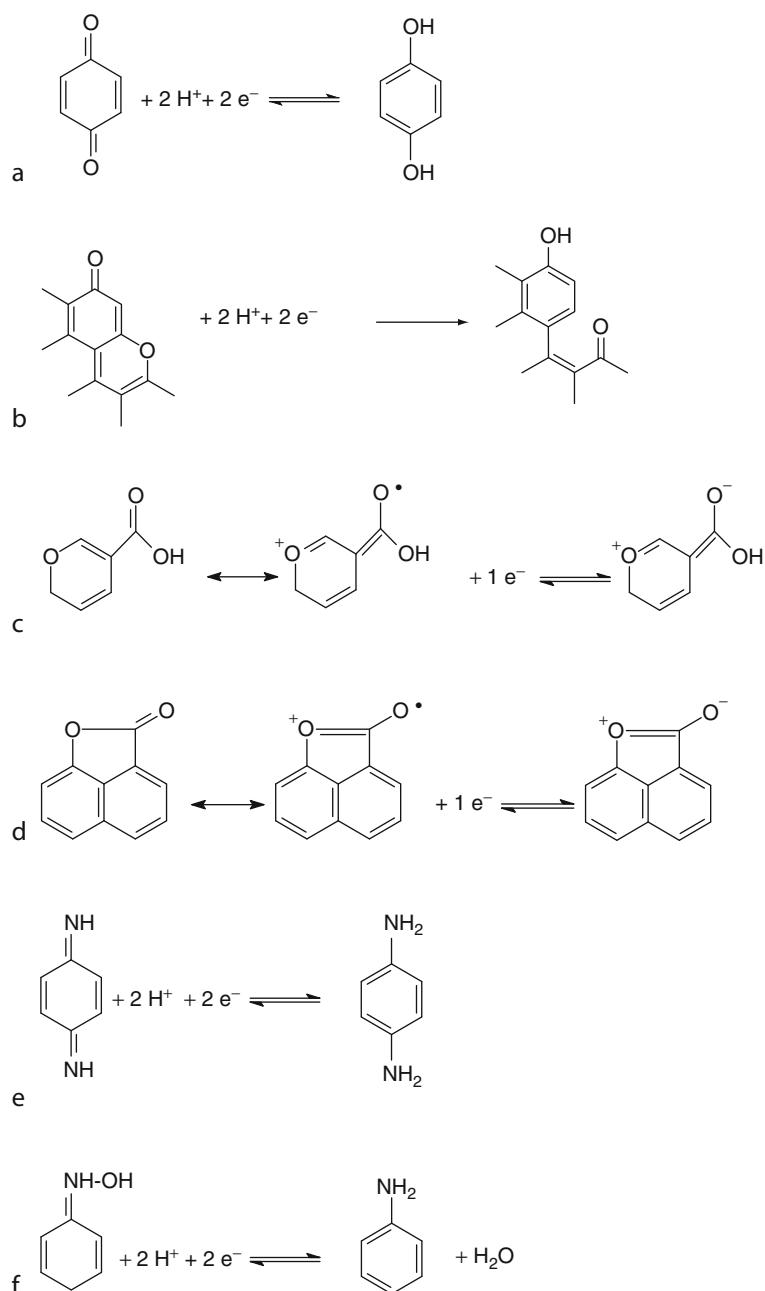
With nanoporous carbon electrodes in organic electrolyte, the double-layer capacitance is the major contribution. By contrast, in protic media (aqueous electrolytes or protic ionic liquids), an important pseudo-capacitive contribution is observed for nanotextured carbons enriched with heteroatoms as nitrogen and oxygen. Reversible hydrogen electrosorption is another kind of pseudo-faradic effect which might be also observed with nanoporous carbons in aqueous medium. In this case, nascent hydrogen formed by water reduction under negative polarization of carbon is adsorbed in the material; polarization reversal provokes its desorption by oxidation. The different types of pseudo-capacitive

contributions for nanoporous carbons are reviewed in the following part of this chapter.

Pseudo-capacitance of Heteroatom-Enriched Carbons

The most reported dopants which confer pseudo-capacitive properties to carbons are oxygen (Fig. 3) and nitrogen (Fig. 4). Some other studies described below have recently considered boron or phosphorous as pseudo-capacitive dopants. The nitrogenated and oxygenated functionalities can undergo pseudo-faradic reactions, which can be pH-dependent or not, as presented in Fig. 10. The extent of the thereof extracted pseudocapacitance depends on the number of surface groups, their chemical nature, and accessibility.

Pseudocapacitance due to oxygenated surface functionalities is the most studied. In particular, the quinone/hydroquinone redox pair involves the exchange of one electron per oxygenated group (see Fig. 10a), against 0.18 electron stored per carbon atom in the electrical double-layer [14]. Other functionalities, such as pyrone-like structures (combination of non-neighboring carbonyl and ether oxygen atoms at the edges of the graphene layers – Fig. 10b), have been shown to be electrochemically active in the same potential range as the quinone/hydroquinone pair [26].



Nanocarbons for Supercapacitors. Figure 10

Possible redox reactions involving oxygenated or nitrogenated surface functionalities

The activity of other functionalities also present in the complex chemistry of a carbon surface is not clear, even if some of them as the carboxylic or lactones could show resonance with other functional groups able to undergo electron transfer reactions (see Fig. 10c, d).

According to the equations in Fig. 10a and b, the electrochemical activity of oxygenated functionalities such as quinones or pyrones depends on the electrolyte pH [27]. The quinone/hydroquinone pair is undoubtedly active at low pH values [27, 28], and it has been claimed that the pyrone-like one is more active at high

than at low pH [27]. Andreas et al. also suggested other unidentified pseudocapacitive reactions occurring on the carbon surface at basic pH. However, any oxygenated surface functionality being electrochemically active at neutral pH has been described. Due to this fact, and taking into account that smaller capacitance values are obtained at neutral pH than at acidic or basic ones [27], the main research efforts on the pseudocapacitive effect of the oxygenated surface functionalities have been focused on using KOH or H₂SO₄ as electrolytes.

Pseudo-capacitance in Basic and Acidic Media

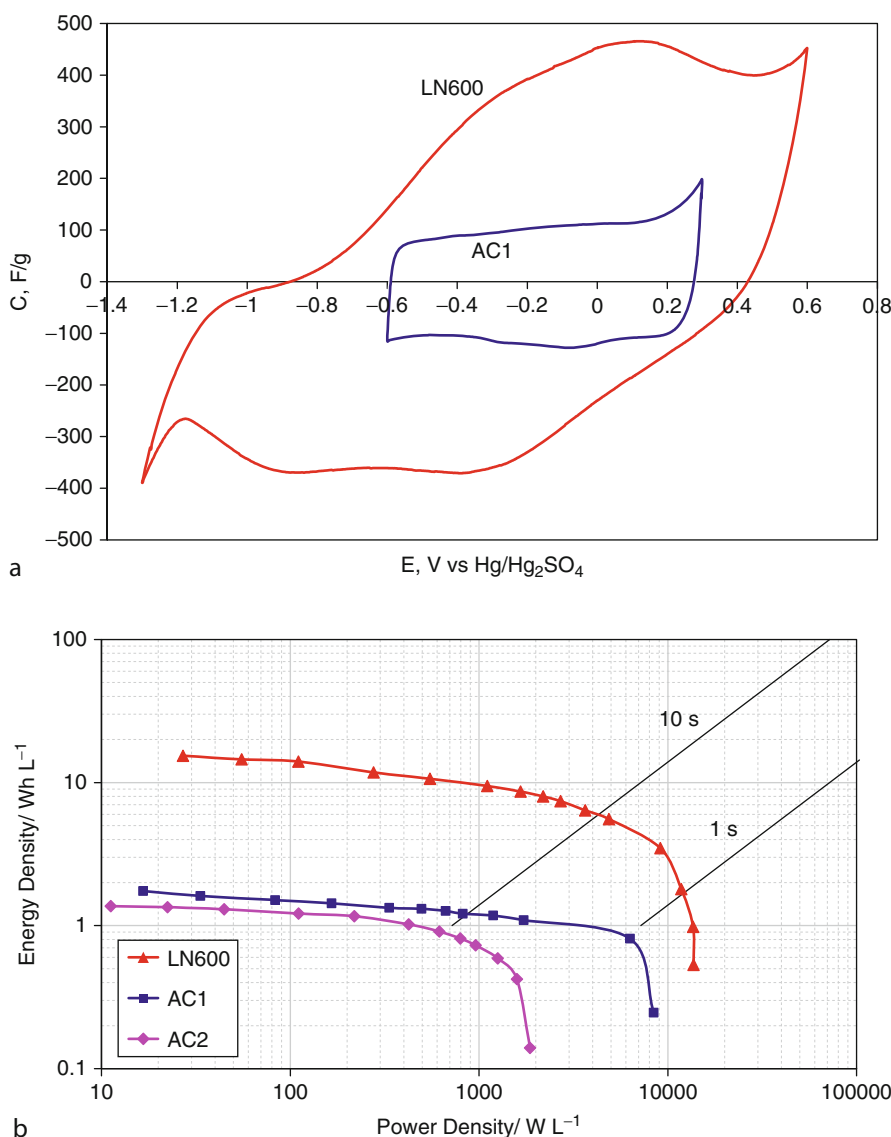
Pseudo-capacitance due to Oxygenated Functionalities The capacitance enhancement of O-enriched carbons in KOH electrolyte has been correlated with the oxygenated functionalities desorbing as CO during a thermal treatment in inert atmosphere, e.g., phenol, ether, carbonyl or quinone groups [29, 30]. A similar correlation between capacitance and the CO-type oxygenated groups has been shown for a series of activated carbons tested in H₂SO₄ [31]. By contrast, the functionalities desorbing as CO₂, e.g., carboxyl, anhydride, or lactone [30], are supposed to hinder the electrolyte migration in the micropores, thus leading to a delay in the electrochemical response.

Although much work is still necessary for determining the exact electrochemical behavior of the individual oxygenated surface functionalities, their contribution in enhancing capacitance of nanoporous carbons in acidic or basic media is a demonstrated fact. However, one of the drawbacks of most nanoporous carbons is their low density which leads to a low volumetric capacitance and consequently to insufficiently compact supercapacitors. Such disadvantage can be circumvented by producing nanotextured carbons through the pyrolysis of seaweed-based biopolymers [32] or seaweeds themselves [33] at mild temperature, e.g., 600°C. The carbons are characterized by a high electrical conductivity and a moderately developed porosity ($S_{\text{BET}}=200\text{--}700\text{ m}^2\text{ g}^{-1}$) and the presence of 10–15 wt% of oxygen. Due to the pseudo-faradic contribution of oxygenated functionalities, the capacitance reaches values as high as 260 F g⁻¹ in 1 mol L⁻¹ H₂SO₄ with these carbons, despite the poorly developed specific surface area. Figure 11a shows three-electrode

cyclic voltammograms of the seaweed-based carbon (LN600) and a commercial-activated carbon AC1 (Norit Super 50; $S_{\text{BET}}=1,400\text{ m}^2\text{ g}^{-1}$; Oxygen=4.6 at.%). AC1 shows an almost rectangular voltammogram, demonstrating that charging the electrical double layer is the dominant contribution, whereas LN600 presents very pronounced and reversible humps, at -0.4 and 0.1 V vs Hg/Hg₂SO₄, related with redox reactions probably involving the quinone and the pyrone-like groups. This surface functionality is suggested to enhance capacitance and to shift the irreversible and destructive oxidation of the material toward more positive potentials. On the negative side of the voltammogram, the potential of dihydrogen evolution is shifted to very negative values (ca. -1.2 V vs Hg/Hg₂SO₄) compared to the thermodynamic potential for water reduction, e.g., -0.62 V vs Hg/Hg₂SO₄ at pH=0. Due to this overpotential, hydrogen is sorbed/inserted in the pores during the cathodic scan (hump at around -0.9 V vs Hg/Hg₂SO₄) and further oxidized (hump at around -0.2 V vs Hg/Hg₂SO₄) during the anodic one, giving rise to an additional pseudo-capacitive contribution [34]. This phenomenon will be discussed in detail in section “Pseudo-capacitance Related with Reversible Hydrogen Electrosorption in Nanoporous Carbons”. As a result of broadening the stability window, a symmetric capacitor built with LN600 electrodes in 1 mol L⁻¹ H₂SO₄ could be charged/discharged at voltage values as high as 1.4 V, whereas only 0.8–1.0 V can be applied with activated carbons [35].

The weakly porous seaweed carbons present a high density, which makes them more interesting than activated carbons in terms of volumetric energy density. Figure 11b shows that the volumetric energy density reached with the carbon LN600 is about one order of magnitude higher than for capacitors based on the commercial-activated carbons AC1 and AC2 (PX21, Kansai, Japan; $S_{\text{BET}}=2,500\text{ m}^2\text{ g}^{-1}$; Oxygen=5.0 at.%).

When seaweeds are carbonized in the presence of multiwalled carbon nanotubes (CNTs), the resulting composites are mesoporous, and the presence of CNTs dramatically enhances their conductivity [36]. Due to this open mesoporosity and the high conductivity of CNTs, the accessibility of ions to the active mass together with the charge propagation is improved. As it can be observed in Fig. 12, for the



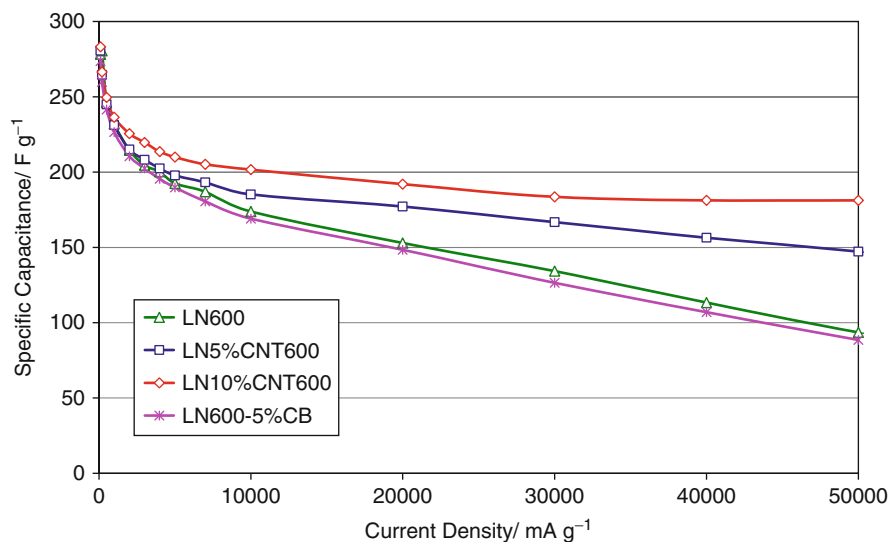
Nanocarbons for Supercapacitors. Figure 11

(a) Three-electrode cyclic voltammograms of LN600- and AC1-based electrodes in 1 mol L⁻¹ H₂SO₄; (b) Volumetric Ragone plots which compare the performance of supercapacitors with LN600-, AC1-, and AC2-based electrodes in 1 mol L⁻¹ H₂SO₄ (Adapted from [33])

composites containing 10 wt% of CNTs, specific capacitance values as high as 180 F g⁻¹ can still be obtained at a current density of 50 A g⁻¹, whereas the addition of carbon black does not give any benefit in comparison to the LN600 material alone. As a consequence, the highest extractable energy at high power density is obtained with CNT-based composite electrodes. In addition, the cycle life of the systems is improved

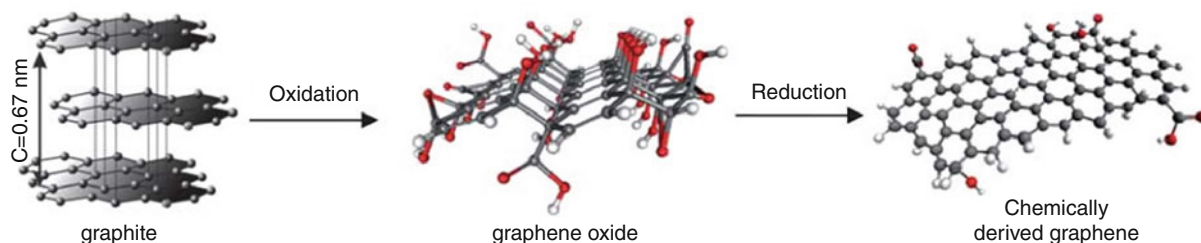
owing to the resiliency introduced by the CNTs in the composite electrodes.

Graphene nanosheets are another kind of moderately developed surface area and highly oxygenated carbon material, which has attracted much attention during the last years [37–41]. In literature, perfectly dispersed individual graphene nanosheets are expected to display a high specific surface area, being promising



Nanocarbons for Supercapacitors. Figure 12

Specific discharge capacitance vs current density for supercapacitors (charged up to 1.2 V) based on LN600 and its composites with 5 wt% CNTs (LN5%CNT600), 10 wt% CNTs (LN10%CNT600), and 5 wt% carbon black (LN600-5%CB) in 1 mol L⁻¹ H₂SO₄ [36]



Nanocarbons for Supercapacitors. Figure 13

Illustration of the chemical route to the synthesis of chemically derived graphene [41]

for EDLC. Unfortunately, in practice, the obtaining of stable single graphene sheets with controllable size and morphology remains a great challenge. Among the different synthesis methods, the chemical one based on graphite oxidation giving graphite oxide, which is subsequently reduced by hydrazine (Fig. 13), is the most proposed because of a possible upscaling, even if harmful chemicals are used. Whatever the preparation method, the resulting material consists in an agglomeration of highly functionalized graphene sheets. As a consequence, the specific surface area is relatively low (300–900 m² g⁻¹) compared to the theoretical one (2,620 m² g⁻¹), and the materials contain 10–20 at.% of oxygen and nitrogen. Obviously, the

capacitance values, from 100 to 200 F g⁻¹, reported for these materials in KOH or H₂SO₄ electrolytes are not related to the nanotexture but mainly to the pseudo-faradic reactions involving the surface functionality at the edges of the graphene sheets. Hence, the literature claims that graphene exhibits good EDL properties are not correct. Moreover, despite the exponential growth of the number of publications on synthesis and applications of graphene for supercapacitors, a cost-effective and environment-friendly mass-production method still remains a challenge [41], and the electrochemical performance should be improved to compete with other existing carbon materials.

Pseudo-capacitance Due to Nitrogenated Functionalities The specific introduction of nitrogen as a dopant of carbon is also extensively reported in the literature. Nitrogen-rich nanocarbons can be obtained either by post-treatment of carbons with N-containing reagents as the NH_3 /air mixture [42] and urea [43] or by pyrolysis of N-containing polymers. Such pyrolysis can be performed on the polymer itself, and followed by an activation step for generating porosity [44–46], or on mixtures of polymer and nanotexturing agent as CNTs [47, 48], zeolites [49], or mica [50].

A correlation between capacitance in aqueous KOH or H_2SO_4 electrolytes and the total amount of surface nitrogen has been shown by Frackowiak et al. [45] for a series of N-doped carbons of similar nanoporous texture. The carbons were obtained by pyrolysis of nitrogen-containing precursors as polyvinylpyridine or polyacrylonitrile and their blends with coal tar pitch, followed by steam activation. Among the various nitrogenated functionalities, it seems that the ones located at the graphene edges, as pyridine and pyrrol (see Fig. 4), are the most active [46, 51]. The presence of nitrogen can also influence differently the capacitance of the positive and negative electrodes of a capacitor [43] due to the potential dependence of the pseudo-faradic reactions.

High-performance nitrogen-rich carbons have been developed by applying the “templating technique.” Briefly a nitrogen-containing precursor is introduced in a nanoporous scaffold and subsequently pyrolyzed; afterward, the nitrogen-enriched carbon replica is obtained by etching the host with hydrofluoric acid. As an example, a weakly porous carbon material ($S_{\text{BET}}=400 \text{ m}^2 \text{ g}^{-1}$) containing 24 at.% of nitrogen was synthesized through melamine intercalation in a lamellar aluminosilicate [50]. The specific capacitance, obtained in three-electrode cell, reaches 200 Fg^{-1} in $1 \text{ mol L}^{-1} \text{ H}_2\text{SO}_4$. Much higher gravimetric capacitance of 340 Fg^{-1} was demonstrated in $1 \text{ mol L}^{-1} \text{ H}_2\text{SO}_4$ by the carbon obtained through pyrolysis of acrylonitrile in NaY zeolite. This carbon displays a higher specific surface area of $1,680 \text{ m}^2 \text{ g}^{-1}$ and a moderate nitrogen amount of 6 at.%. The exceptionally high capacitance comes from the synergy of different contributions: (1) the ordered superstructure inherited from the scaffold which favors the diffusion of ions to the active surface; (2) micropores with a size

perfectly adapted to the electrolyte which provide an EDL contribution; and (3) an important pseudo-capacitive effect related with the surface functionality [49]. Moreover, a symmetric capacitor from these carbons could operate up to 1.2 V during 10,000 charge/discharge cycles, which is much higher than the usual maximum voltage for systems based on activated carbon electrodes in aqueous media [35]. Hence, nitrogen enrichment seems to provide a profitable potential shift of the electrodes as it has been observed for oxygen [33].

Following the industrial requirements, the tendency now is to develop heteroatom-enriched carbons enabling to operate at high current density. For example, carbon/CNTs nanocomposites have been prepared by one-step carbonization of polyacrylonitrile/carbon nanotube blends [47] and melamine-formaldehyde/carbon nanotube blends [48]. Although presenting a moderate specific surface area ($S_{\text{BET}} \sim 400 \text{ m}^2 \text{ g}^{-1}$), the later carbons display gravimetric capacitance values of 170 and 130 Fg^{-1} at current densities of 50 and 5 Ag^{-1} , respectively [48]. Such exceptional performance is due, as for the previously presented LN600/CNTs composites, to the synergetic effects of nanotubes and surface functionality. The open mesoporosity of nanotubes and their good electrical conductivity enable good propagation of charges.

Recently, capacitance values around $100\text{--}200 \text{ Fg}^{-1}$ have been claimed for nonporous (specific surface area ranging from 6 to $86 \text{ m}^2 \text{ g}^{-1}$) N-enriched carbons prepared by ammonia treatment of an already nitrogenated carbon derived from melamine [52]. However, the data were obtained at very low current density and the study does not provide any information on the performance of these materials in usual operating conditions of supercapacitors.

Even if the previously presented results seem to demonstrate the efficiency of nitrogen as carbon dopant for introducing a pseudocapacitive character, one must take into account that oxygen is always present on the surface of carbons materials. Studies performed on model compounds as crystalline oligomers of 1,5-diaminoanthraquinone (DAAQ) with a 3D-extended π -conjugated structure [53, 54] have shown that the redox peaks corresponding to quinone/hydroquinone and emeraldine/quinone diimide pairs arise in the same potential range. Therefore, the capacitance values

normally assessed only to the nitrogenated functionalities may include a contribution due to the presence of oxygenated surface functionalities.

Conclusion O- and N-enriched nanocarbons demonstrate high capacitance values in KOH or H_2SO_4 due to the contribution of pseudo-faradic reactions. Another positive effect of doping is the broadening of the potential stability window. Composites of these carbons with carbon nanotubes withstand remarkable capacitance values at high current load. Hence, the N- and O-doped carbons open the opportunity of developing high-performance supercapacitors in aqueous electrolytes.

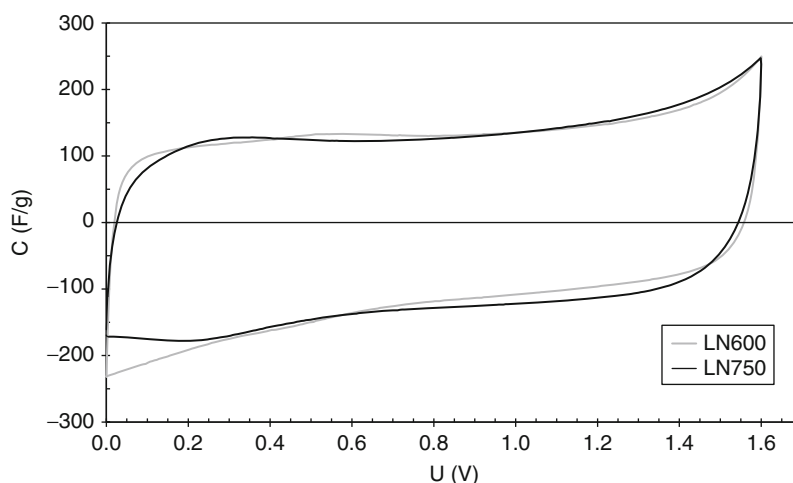
Recently, boron [55] and phosphorous [56] have been tested as dopants for improving the electrochemical performance of carbons in acidic electrolytes. However, since the investigated boron-doped carbons contained also nitrogen and oxygen, the peculiar role of boron is not established. For the particular case of phosphorous, it seems that small amounts affect positively the overall capacitance and improve the capacitance retention ratio at high current load in H_2SO_4 .

Pseudo-capacitance in Neutral Medium Neutral electrolytes have been scarcely applied for carbon-based supercapacitors. Indeed, the capacitance values of activated carbons in neutral electrolytes are lower than in other aqueous electrolytes, suggesting that the

pseudo-faradic redox reactions of O and N electroactive surface groups are depressed in comparison with acidic or basic electrolytes [25, 27, 57–59]. Therefore, it was concluded that oxygenated or nitrogenated functionalities are electrochemically inactive in neutral electrolytes and that only the formation of the electrical double layer contributes to the capacitance in this media.

Nevertheless, a recent study has shown that a contribution of pseudo-faradic reactions is possible in a neutral electrolyte, e.g., Na_2SO_4 , if the carbon material possesses an adequate surface functionality [60]. Figure 14 shows the cyclic voltammograms of supercapacitors built with nanotextured carbons obtained after pyrolysis of seaweeds at 600°C (LN600, $S_{\text{BET}}=746 \text{ m}^2 \text{ g}^{-1}$) and at 750°C (LN750, $S_{\text{BET}}=1,082 \text{ m}^2 \text{ g}^{-1}$). Despite the lower specific surface area of LN600 in comparison to LN750, the capacitive behavior is comparable for the two carbons. Taking into account that LN600 and LN750 contain 11.5 at.% and 7.3 at.% of oxygen, respectively, one can conclude that pseudo-faradic reactions contribute more markedly to the capacitance of LN600 in Na_2SO_4 . In particular, the fact that quinone-like groups are twice more present in LN600 than in LN750 suggests that the latter are the most active ones at $\text{pH}=6.4$.

Notwithstanding these facts, the most striking result obtained in neutral electrolytes is the ability of symmetric carbon/carbon supercapacitors to operate



Nanocarbons for Supercapacitors. Figure 14

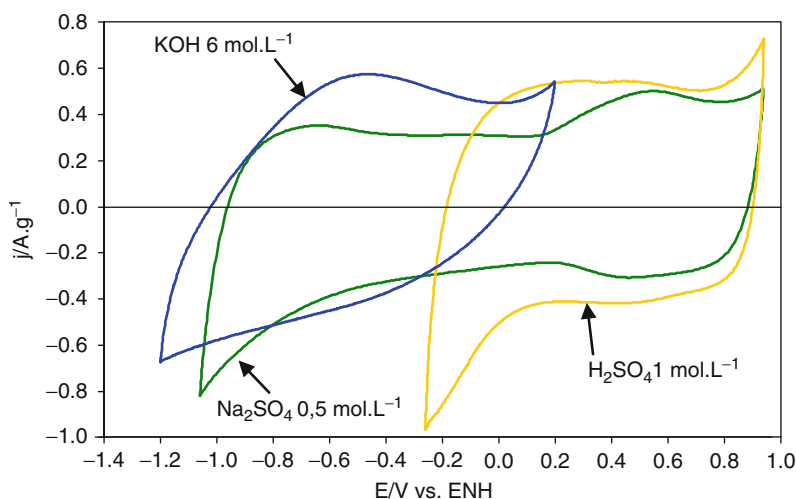
Cyclic voltammograms of LN600- and LN750-based symmetric capacitors in $0.5 \text{ mol L}^{-1} \text{ Na}_2\text{SO}_4$ [60]

up to voltages as high as 1.6 V during more than 10,000 cycles [60, 61]. According to Eq. 5, the energy density is higher in neutral than in other aqueous electrolytes.

The former result can be understood by comparing the three-electrode cyclic voltammograms of a commercial-activated carbon (AC, $S_{\text{BET}} = 2,250 \text{ m}^2 \text{ g}^{-1}$, O = 2.5 at.%; provided by MeadWestvaco, USA) in acidic (H_2SO_4), basic (KOH), and neutral (Na_2SO_4) electrolytes (Fig. 15) [61]. In Na_2SO_4 , the stability potential window is about twice larger than in H_2SO_4 and KOH. Actually, in the neutral electrolyte, it reaches 2.0 V for AC and values of even 2.4 V have been found for seaweed-based carbons [60]. The reason of such a wide stability potential window is the high overpotential for dihydrogen evolution, e.g., 0.6 V, and the better oxidation resistance of carbons in the neutral medium. A pH dependence of the overpotential for H_2 evolution has been also shown with glassy carbon [62] or Pt [63] electrodes. In a neutral pH solution, the overpotential reaches 1.2 V with a glassy carbon electrode, whereas only values of 0.6 V or 0.8 V are attained in acidic or basic pH, respectively. With a Pt electrode, a small overpotential of 0.3 V is found in neutral electrolyte, whereas any overpotential can be detected in acidic or basic solutions. For the particular case of Pt [63], this property is attributed to the fact that the H_3O^+ concentration in neutral electrolyte is

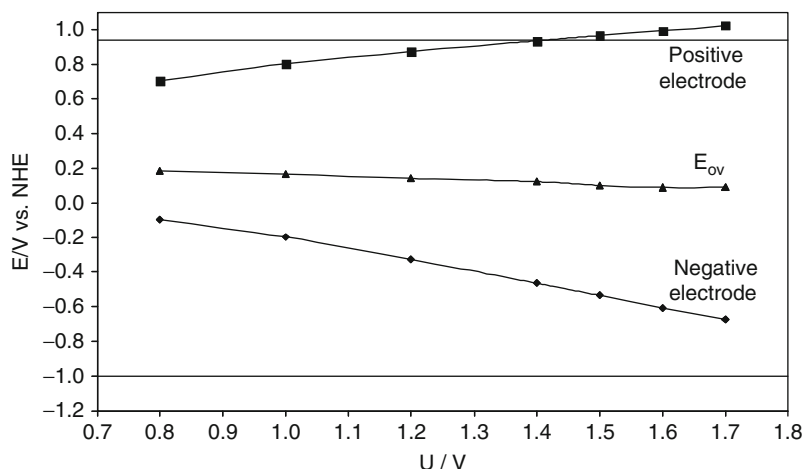
too low to induce gas evolution reactions. For nanoporous carbons, it is also related with their storage ability of nascent hydrogen below the thermodynamic potential for water reduction. This pseudo-capacitive contribution related to hydrogen sorption will be discussed in detail in the section “Pseudo-capacitance Related with Reversible Hydrogen Electrosorption in Nanoporous Carbons”.

Although 1.6 V is the highest voltage ever claimed for a symmetric carbon/carbon capacitor in an aqueous electrolyte, this value is smaller than the stability potential window of 2.0 V determined for AC in three-electrode cell. The reasons of the potential window underuse in Na_2SO_4 have been demonstrated after experiments using a special two-electrode cell equipped with a reference electrode. Such configuration allows the positive and negative electrode potentials to be recorded separately during cycling the supercapacitor between zero and a given maximum voltage. Figure 16 shows that, if the supercapacitor is charged between 0 V and the maximum stability voltage of 1.6 V, the operating potential window for the negative and the positive electrode is [0.09; -0.61] and [0.09; 0.99] V vs NHE, respectively. According to Fig. 15, in Na_2SO_4 electrolyte, the lowest potential for a negative electrode before dihydrogen production is around -1 V vs NHE and the highest one for avoiding an irreversible



Nanocarbons for Supercapacitors. Figure 15

Three-electrode cyclic voltammograms showing the potential stability window of AC in 6 mol L^{-1} KOH, 1 mol L^{-1} H_2SO_4 , and 0.5 mol L^{-1} Na_2SO_4 [61]



Nanocarbons for Supercapacitors. Figure 16

Potential reached by the positive and negative electrodes during the operation of a 2-electrode cell (equipped with a reference electrode) at different values of maximum voltage. The E_{ov} values correspond to the electrodes potential when the cell voltage U is set to 0 V. Electrolyte: $0.5 \text{ mol L}^{-1} \text{ Na}_2\text{SO}_4$. The horizontal lines represent the extreme potential values determined for AC in 3-electrode cell [61]

oxidation of a positive electrode is around 0.95 V vs NHE (see the horizontal lines included in Fig. 16). Hence, Fig. 16 clearly shows that the maximum voltage of the supercapacitor is limited by the positive electrode. In fact, the maximum potential reached by the positive electrode at $U=1.6 \text{ V}$ is slightly beyond the oxidation limit, whereas the minimum potential reached by the negative electrode is still far away, ca. 0.4 V, from the limit imposed by dihydrogen evolution. During cycling, the maximum potential of the positive electrode slightly shifts to lower values, allowing a good cyclability of the capacitor.

From the foregoing, even if the AC/AC symmetric capacitor in Na_2SO_4 largely outperforms all the symmetric systems in other aqueous electrolytes, the strategy to follow for taking advantage of the whole stability potential window of 2.0 V, and for enhancing the energy density, would consist to broadening the potential range of the negative electrode. In sulfuric acid medium, it has been shown that this kind of shift of the electrodes potential range is possible by building asymmetric systems combining two different carbon materials with a surface functionality adapted for optimally operating as positive or negative electrode [64].

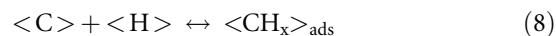
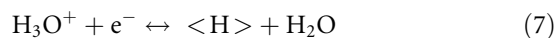
In conclusion, the nature of the surface functionalities of carbons and the electrolyte pH influence both the capacitance values and stability potential window.

Actually, the lately developed supercapacitors in neutral medium are able to compete with the organic electrolyte-based systems while being environment friendly.

Pseudo-capacitance Related with Reversible Hydrogen Electrosorption in Nanoporous Carbons

As it has been introduced in the previous section, hydrogen electrosorption in carbon materials under negative polarization has a direct impact on the energy density of a supercapacitor operating in an aqueous electrolyte. Due to the overpotential of dihydrogen evolution, the electrochemical stability window can be extended to lower potential values. Moreover, the electro-desorption of hydrogen by anodic oxidation gives rise to a pseudo-faradic contribution in addition to the EDL capacitance of the material.

Under negative polarization of a nanotextured carbon electrode in aqueous electrolyte, hydrated cations are first adsorbed forming an electrical double layer. When the electrode potential is lower than the equilibrium value for water reduction, nascent hydrogen is formed (Eq. 7) and adsorbed onto the carbon surface [34] (Eq. 8):

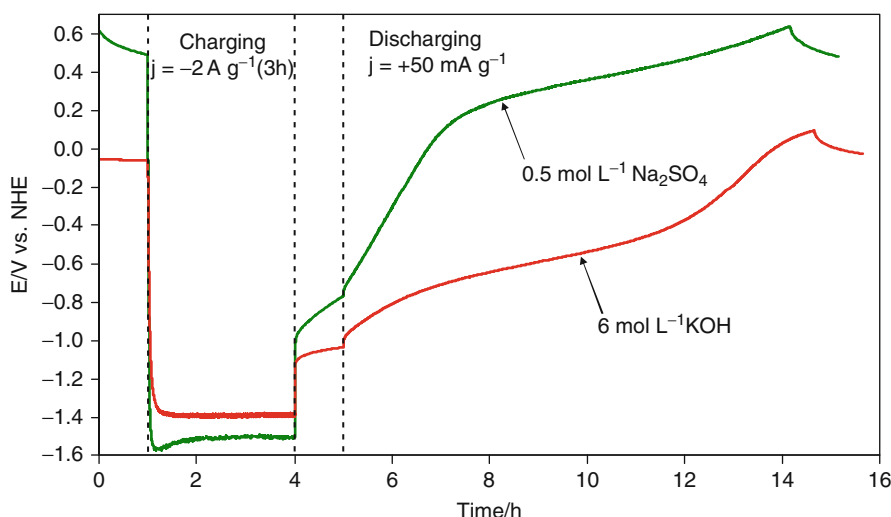


where $\langle H \rangle$ represents the nascent hydrogen, $\langle C \rangle$ the carbon host, and $\langle CH_x \rangle$ hydrogen adsorbed in the later.

It has been demonstrated that the nanoporous texture of the activated carbon plays an important role [22, 34, 65–67]. In particular for carbons prepared by a templating technique, the amount of sorbed hydrogen is proportional to the ultramicropore (0.6–0.7 nm) volume [22]. In this case, straight mesopores are very useful channels for the transport of ions. Galvanostatic charge/discharge experiments in alkaline medium on a series of activated carbons with similar porosity but with different surface chemistry show that the amount of hydrogen reversibly electrosorbed decreases with the increase of the oxygenated surface functionalities [68]. The electrosorption of hydrogen has been extensively studied in alkaline or acidic medium [34]. In such media, the important polarization between reduction and oxidation potentials suggests that the interaction of sorbed hydrogen with the carbon substrate is stronger than physisorption. In the thermo programmed desorption (TPD) analyses realized on carbon samples submitted to galvanostatic charge in KOH medium, a hydrogen peak appears at around 200°C [69]. From this peak, the activation energy was estimated

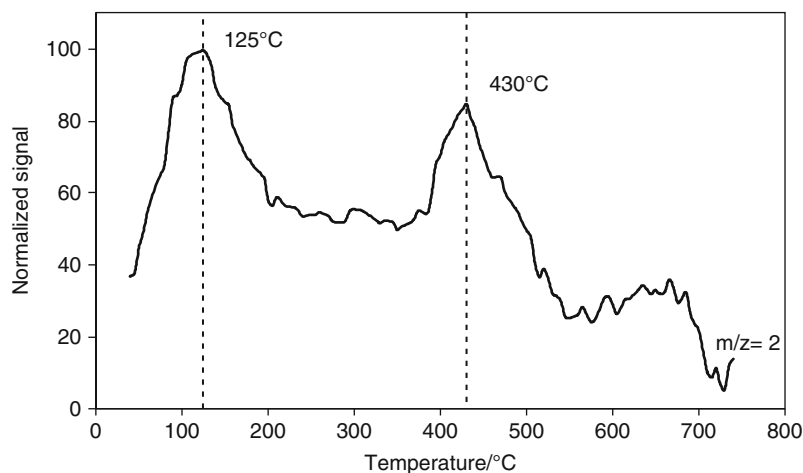
to 110 kJ mol^{−1} for hydrogen desorption, confirming a weakly chemisorbed state of hydrogen.

A recent study demonstrates the effectiveness of a neutral medium for electrochemical hydrogen bonding to the graphene units in nanocarbons [70]. From the galvanostatic charge/discharge curves (Fig. 17), the mechanism for hydrogen sorption seems different in neutral medium and in KOH. In KOH, a single plateau at ca 0.6 V vs NHE characterizes the anodic oxidation of hydrogen [34]. By contrast, in neutral medium (0.5 mol L^{−1} Na₂SO₄), the curve exhibits two different slopes, a first part where the stored hydrogen is oxidized at potentials lower than 0.1 V vs NHE, and a plateau at higher potentials corresponding to more strongly bonded hydrogen. It has also to be noted that the reduction overpotential is more important in neutral than in alkaline medium. The TPD analysis realized after the galvanostatic charge of AC in the neutral medium (Fig. 18) confirms two types of hydrogen bonding with peaks at 125°C and 430°C. The desorption activation energies calculated from these data are 110 and 205 kJ mol^{−1}, respectively. The highest value demonstrates that the main part of nascent hydrogen is strongly bonded to the graphene units in 0.5 mol L^{−1} Na₂SO₄.



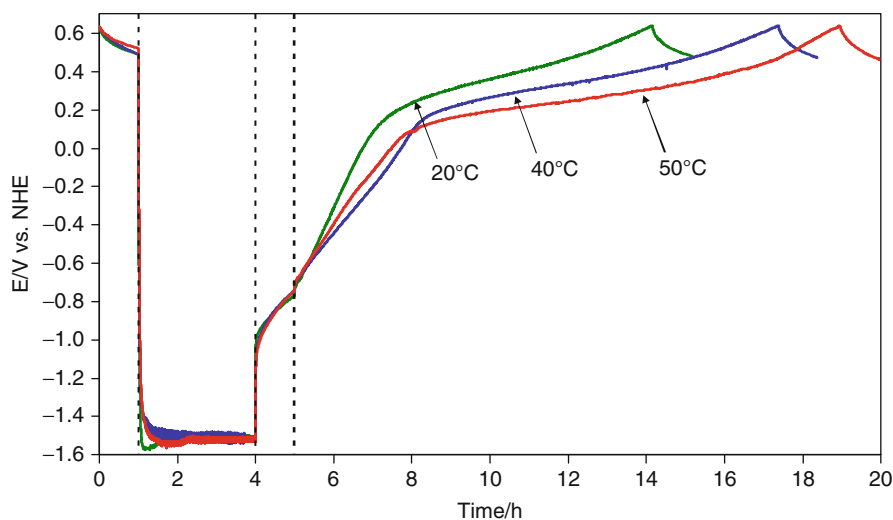
Nanocarbons for Supercapacitors. Figure 17

Galvanostatic charge/discharge of AC in 6 mol L^{−1} KOH (red curve) and 0.5 mol L^{−1} Na₂SO₄ (green curve). After 1 h relaxation, a negative current of −2 A g^{−1} is applied during 3 h. Then, after 1 h relaxation, a positive current of +50 mA g^{−1} is applied in order to oxidize the sorbed hydrogen [70]



Nanocarbons for Supercapacitors. Figure 18

TPD analysis of dihydrogen evolved from AC power charged at -2 A g^{-1} during 3 h at 50°C in $0.5 \text{ mol L}^{-1} \text{ Na}_2\text{SO}_4$ [70]



Nanocarbons for Supercapacitors. Figure 19

Galvanostatic charge/discharge of AC in $0.5 \text{ mol L}^{-1} \text{ Na}_2\text{SO}_4$ at 20°C , 40°C , and 50°C [70]

The effect of temperature on the electrochemical performance in $0.5 \text{ mol L}^{-1} \text{ Na}_2\text{SO}_4$ has been evaluated (Fig. 19). The oxidation time of stored hydrogen increases with temperature (1.5 wt% at 20°C , 2.1 wt% at 40°C and 2.4 wt% at 50°C), confirming a weakly chemisorbed state of hydrogen [69]. Meanwhile, the oxidation potential decreases when temperature increases, due to a better conductivity of the electrolytic medium. Hence, the energy density of a cell using activated carbon as negative electrode in neutral

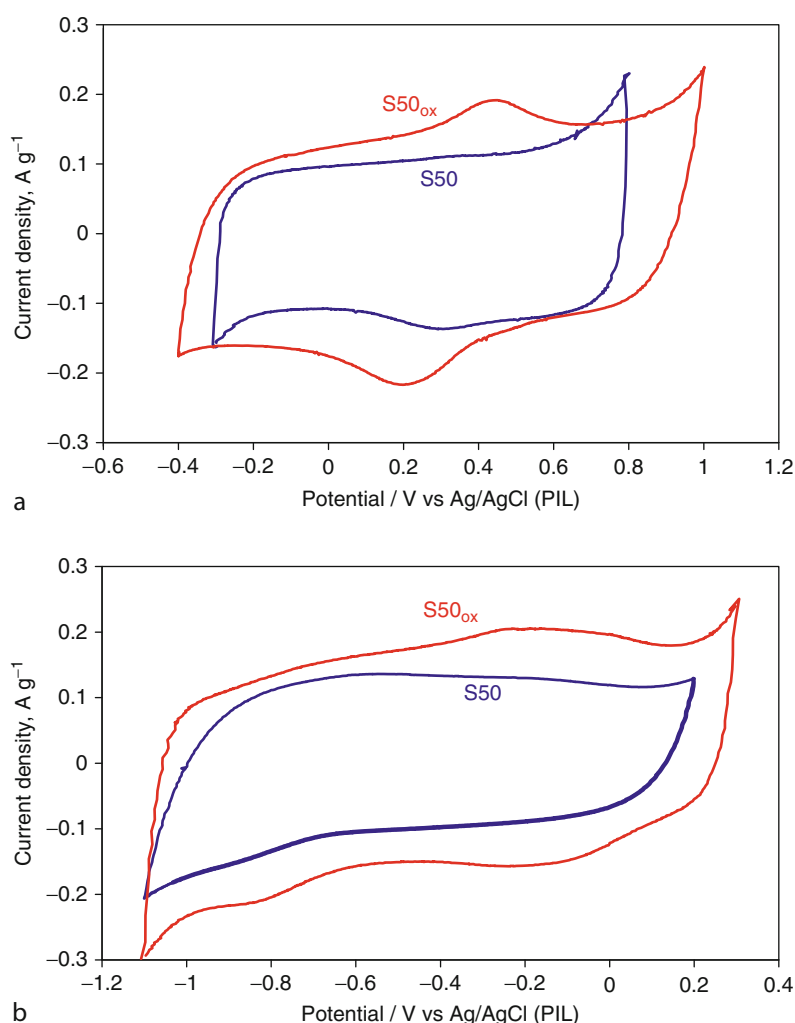
aqueous solution will be enhanced by operating at higher values of temperature.

To summarize, in comparison to acidic and basic electrolytes, a neutral medium has two positive impacts for supercapacitor applications: (1) the overpotential for dihydrogen evolution is higher, allowing the stability window to be enlarged; (2) the amount of hydrogen reversibly sorbed is enhanced enabling a higher pseudo-capacitive contribution related with its oxidation.

Protic Ionic Liquids as Electrolytes for Carbon-Based Pseudo-capacitors

Protic ionic liquids (PILs) are relatively cheap and environment friendly electrolytes [71]. Due to their proton-exchanging capabilities, a pseudo-faradaic enhancement of capacitance can be expected, as in aqueous electrolytes, while operating in a larger potential window. For these reasons, they have been proposed as electrolytes in carbon-based supercapacitors [72]. Cyclic voltammograms in Fig. 20 show the electrochemical performance of two pyrrolidinium-based

PILs (Pyrrolidinium nitrate [Pyr][NO₃] (IL1) and pyrrolidinium formate [Pyr][HCOO] (IL2)) [73] in presence of two different carbon electrodes: (1) a commercial-activated carbon (S50 from Norit, $S_{\text{BET}}=1,400 \text{ m}^2 \text{ g}^{-1}$, O=4.6 at.%); (2) the same oxidized in order to increase the amount of surface functionalities without modifying the porous texture (S50_{ox}, O=9.7 at.%). For IL1 (Fig. 20a), redox peaks are seen at around 0.2 and 0.45 V vs Ag/AgCl(PIL). For activated carbons, peaks in such potential range are traditionally assessed to the electrochemical reactions



Nanocarbons for Supercapacitors. Figure 20

Three-electrode cyclic voltammograms with activated carbon (S50) or oxidized activated carbon (S50_{ox})-based electrodes in (a) pyrrolidinium nitrate and (b) pyrrolidinium formate vs the Ag/AgCl(sat, KCl(sat (PIL) reference electrode (Adapted from [71])

of oxygenated surface functionalities such as quinone/hydroquinone pairs [14] or pyrone-like structures able to accept two protons and two electrons [26]. For both ILs, the most pronounced redox peaks are observed for S50_{ox} (Fig. 20a, b), that is evidently caused by a stronger pseudo-capacitive effect linked to the higher amount of quinone-type groups and/or pyrone in S50_{ox}.

Two-electrode cells based on IL1 and IL2 with the two carbons S50 and S50_{ox} are able to operate up to 1.2 V, which is a substantially higher value than the maximum of 0.7 – 1 V generally observed in aqueous electrolytes [35]. Distinct improvements in capacitance owing to the use of S50_{ox} are noticeable for both ionic liquids; the capacitance increases from 99 to 126 Fg⁻¹ in IL1 and from 96 to 130 Fg⁻¹ in IL2, when using S50 with S50_{ox}, respectively. These values are comparable to those obtained in 1 mol.L⁻¹ H₂SO₄ aqueous solution, e.g., 101 Fg⁻¹ for S50 and 137 Fg⁻¹ for S50_{ox}, demonstrating a remarkable pseudo-faradic efficiency of these ionic liquids.

Hence, protic ionic liquids represent a new kind of green electrolyte for developing high-performance capacitors involving both electrical double-layer charging and a pseudo-faradic mechanism. The electrochemical stability and conductivity of protic ILs, the relation between the carbon pore size and the electrolyte ion size, and the design of electrochemical cells are the most important matters for thorough investigations.

Conclusion and Future Directions

Nanocarbons are the most widely used material for supercapacitor electrodes because, among other characteristics, they are cheap, chemically inert, electrically conductive, and their nanoporous texture and surface functionality are highly versatile.

In the case of capacitors in organic electrolyte, it has been demonstrated that the electrical double-layer (EDL) properties are strongly influenced by the pore size. With the tetraethylammonium tetrafluoroborate salt, the largest values of normalized capacitance are observed for pores smaller than 1 nm, with a maximum at around 0.7–0.8 nm, demonstrating that ions are essentially desolvated in the pores. In situ or ex situ techniques, such as small angle neutron scattering or nuclear magnetic resonance, should be applied for better understanding how ions occupy porosity during

the charge/discharge of electrodes. Such investigations would help to better defining the optimal porous texture. Beside, taking into account that saturation of porosity can occur with carbons developing essentially subnanometric pores, new processes have to be investigated for developing the surface area of carbons while keeping only such pores. Finally, in light of some literature claims that the EDL properties are not only determined by the porous texture, another research direction should consider the structural parameters of carbons.

In protic media (aqueous electrolytes or protic ionic liquids) a pseudo-faradic mechanism can arise together with the formation of the double layer. In this case, the capacitance and the stability potential window are influenced both by the electrolyte pH and the nature of carbon functionalities. In the future, fundamental studies should be realized on carbons with controlled functionalities, using for example grafting techniques, in order to determine which functionalities are the most effective for pseudo-faradic processes. The demonstration that neutral aqueous electrolytes can offer capacitors with operating voltage values as high as 2 V is an important breakthrough of the recent years. The reasons for the important overpotential of dihydrogen evolution in these media must be elucidated, that could allow developing even higher voltage values. Different neutral electrolytes must also be more systematically screened versus carbons with various functionalities in order to enhance capacitance, electrical conductivity, and the operating temperature range. In this strategy, protic ionic liquids might be a new class of electrolytes opening new insights for optimizing electrochemical capacitors.

In summary, the high versatility of nanocarbons in terms of texture and surface functionality allows to perfectly adapt them to the different mechanisms depending on the electrolyte. Such versatility still offers high potentialities for developing high-performance supercapacitors.

Bibliography

1. Marsh H, Rodríguez-Reinoso F (2006) Activated carbons. Elsevier, London
2. Lillo-Ródenas MA, Juan-Juan J, Cazorla-Amorós D, Linares-Solano A (2004) Carbon 42:1371–1375
3. Raymundo-Piñero E, Azais P, Cacciaguerra T, Cazorla-Amorós D, Linares-Solano A, Béguin F (2005) Carbon 43:786–795

4. Jagtoyen M, Derbyshire F (1998) *Carbon* 36:1085–1097
5. Bandosz TJ, Biggs MJ, Gubbins KE, Hattori Y, Pikunic J, Thomson K (2003) Molecular models of porous carbons. In: Radovic LR (ed) *Chemistry and physics of carbon*. Marcel Dekker, New York, pp 137–199
6. Pikunic J, Gubbins KE, Pellenq RJM, Cohaut N, Rannou I, Guet JM, Clinard C, Rouzaud JN (2002) *Appl Surf Sci* 196:98–104
7. Harris PJF, Tsang SC (1997) *Philos Mag* A76:667–677
8. Harris PJF (1997) *Inter Mater Rev* 42:206–218
9. Harris PJF (2005) *Critical Rev Solid State Mater Sci* 30:235–253
10. Radovic LR, Bockrath B (2005) *J Am Chem Soc* 127:5917–5927
11. Azaïs P, Duclaux L, Florian P, Massiot D, Lillo-Ródenas MA, Linares-Solano A, Peres JP, Jehoulet C, Béguin F (2007) *J Power Sourc* 171:1046–1053
12. Pels JR, Kapteijn F, Moulijn JA, Zhu Q, Thomas KM (1995) - *Carbon* 33:1641–1653
13. Kapteijn F, Moulijn JA, Matzner S, Boehm HP (1999) *Carbon* 37:1143–1150
14. Conway BE (1999) *Electrochemical supercapacitors – scientific fundamentals and technological applications*. Kluwer/Plenum, New York
15. Kötz R, Carlen M (2000) *Electrochimica Acta* 45:2483–2498
16. Raymundo-Piñero E, Kierzek K, Machnikowski J, Béguin F (2006) *Carbon* 44:2498–2507
17. Gregg SJ, Sing KSW (1982) *Adsorption, surface area and porosity*. Academic, London, pp 103–104
18. Kaneko K, Ishii C (1992) *Colloid Surf* 67:203–212
19. Barbieri O, Hahn M, Herzog A, Kötz R (2005) *Carbon* 43:1303
20. Chmiola J, Yushin G, Gogotsi Y, Portet C, Simon P, Taberna PL (2006) *Science* 313:1760–1763
21. Chmiola J, Largeot C, Taberna PL, Simon P, Gogotsi Y (2008) *Angew Chem* 120:3440–3443
22. Vix-Guterl C, Frackowiak E, Jurewicz K, Friebe M, Parmentier J, Béguin F (2005) *Carbon* 43:1293–1302
23. Mysyk R, Raymundo-Piñero E, Béguin F (2009) *Electrochem Commun* 11:554–556
24. Ruch PW, Hahn M, Cericola D, Menzel A, Kötz R, Wokaun A (2010) *Carbon* 48:1880–1888
25. Frackowiak E, Béguin F (2001) *Carbon* 39:937
26. Montes-Moran MA, Suarez D, Menendez JA, Fuente E (2004) *Carbon* 42:1219–1225
27. Andreas HA, Conway BE (2006) *Electrochimica Acta* 51:6510–6520
28. Okajima K, Ohta K, Sudoh M (2005) *Electrochimica Acta* 50:2227–2231
29. Cheng PZ, Teng H (2003) *Carbon* 41:2057–2063
30. Figueiredo JL, Pereira MFR, Freitas MMA, Orfao JJM (1999) - *Carbon* 37:1379–1389
31. Bleda-Martínez MJ, Maciá-Agulló JA, Lozano-Castelló D, Morallón E, Cazorla-Amorós D, Linares-Solano A (2005) *Carbon* 43:2677–2684
32. Raymundo-Piñero E, Leroux F, Béguin F (2006) *Adv Mater* 18:1877–1882
33. Raymundo-Piñero E, Cadek M, Béguin F (2009) *Adv Funct Mater* 19:1–8
34. Jurewicz K, Frackowiak E, Béguin F (2004) *Appl Phys A* 78:981–985
35. Ruiz V, Santamaría R, Granda M, Blanco C (2009) *Electrochimica Acta* 54:4481–4486
36. Raymundo-Piñero E, Cadek M, Wachtler M, Béguin F (2011) *ChemSusChem*. doi:10.1002/cssc.201000376
37. Stoller MD, Park S, Zhu Y, An J, Ruoff RS (2008) *Nano Lett* 8:3498–3502
38. Vivekchand SRC, Rout CS, Subrahmanyam KS, Govindaraj A, Rao CNR (2008) *J Chem Sci* 20:9–13
39. Wang Y, Shi Z, Huang Y, Ma Y, Wang C, Chen M, Chen Y (2009) *J Phys Chem C* 113:13103–13107
40. Du X, Guo P, Song H, Chen X (2010) *Electrochimica Acta* 55:4812–4819
41. Zhang LL, Zhou R, Zhao XS (2010) *J Mater Chem* 20:5983–5992
42. Jurewicz K, Babel K, Ziolkowski A, Wachowska H (2003) *Electrochimica Acta* 48:1491–1498
43. Jurewicz K, Pietrzak R, Nowicki P, Wachowska H (2008) *Electrochimica Acta* 53:5469–5475
44. Lota G, Grzyb B, Machnikowska H, Machnikowski J, Frackowiak E (2005) *Chem Phys Lett* 404:53–58
45. Frackowiak E, Lota G, Machnikowski J, Vix-Guterl C, Béguin F (2006) *Electrochimica Acta* 51:2209–2214
46. Ra EJ, Raymundo-Piñero E, Lee YH, Béguin F (2009) *Carbon* 47:2984–2992
47. Béguin F, Szostak K, Lota G, Frackowiak E (2005) *Adv Mat* 17:2380–2384
48. Lota G, Lota K, Frackowiak E (2007) *Electrochem Commun* 9:1828–1832
49. Ania CO, Khomenko V, Raymundo-Piñero E, Parra JB, Béguin F (2007) *Adv Funct Mater* 17:1828–1836
50. Hulicova D, Kodama M, Hatori H (2006) *Chem Mater* 18:2318–2326
51. Hulicova-Jurcakova D, Seredych M, Lu GQ, Bandosz TJ (2009) *Adv Funct Mater* 19:438–447
52. Hulicova-Jurcakova D, Kodama M, Shiraishi S, Hatori H, Zhu ZH, Lu GQ (2009) *Adv Funct Mater* 19:1800–1809
53. Naoi K, Suematsu S, Hanada M, Takenouchi HJ (2002) *J Electrochem Soc* 149:472–476
54. Gao M, Yang F, Wang X, Zhang G, Liu L (2007) *J Phys Chem C* 111:17268–17274
55. Konno H, Ito T, Ushiro M, Fushimi K, Azumi K (2010) *J Power Sourc* 195:1739–1746
56. Hulicova-Jurcakova D, Puziy AM, Poddubnaya OI, Suarez-García F, Tascon JMD, Lu GQ (2009) *J Am Chem Soc* 131:5026–5027
57. Hu CC, Wang CC, Wu FC, Tseng RL (2007) *Electrochimica Acta* 52:2498–2505
58. Subramanian V, Luo C, Stephan AM, Nahm KS, Thomas S, Wei B (2007) *J Phys Chem C* 111:7527–7531
59. Qu QT, Wang B, Yang LC, Shi Y, Tian S, Wu YP (2008) *Electrochem Commun* 10:1652–1655
60. Bichat MP, Raymundo-Piñero E, Béguin F (2010) *Carbon* 48:4351–4361
61. Demarconnay L, Raymundo-Piñero E, Béguin F (2010) *Electrochem Commun* 12:1275–1278

62. Wang J, Kirgoz UA, Mo JW, Lu J, Kawde AN, Muck A (2001) *Electrochem Comm* 3:203–208
63. Hong MS, Lee SH, Kim SW (2002) *Electrochem Solid State Lett* 5:A227–A230
64. Khomenko V, Raymundo-Piñero E, Béguin F (2010) *J Power Sourc* 195:4234–4241
65. Jurewicz K, Frackowiak E, Béguin F (2001) *Electrochem Solid State Lett* 4:A27–A29
66. Jurewicz K, Frackowiak E, Béguin F (2002) *Fuel Process Technol* 77–78:415–421
67. Fang B, Zhou H, Honma I (2006) *J Phys Chem B* 110:4875–4880
68. Bleda-Martínez MJ, Pérez JM, Linares-Solano A, Morallón E, Cazorla-Amorós D (2008) *Carbon* 46:1053–1059
69. Béguin F, Friebe M, Jurewicz K, Vix-Guterl C, Dentzer J, Frackowiak E (2006) *Carbon* 44:2392–2398
70. Demarconnay L, Raymundo-Piñero E, Béguin F (2010) *CD Proceedings of the International Carbon Conference, Clemson, USA, 2010*
71. Greaves TL, Drummond CJ (2008) *Chem Rev* 108:206–223
72. Mysyk R, Raymundo-Piñero E, Anouti M, Lemordant D, Béguin F (2010) *Electrochem Comm* 12:414–417
73. Anouti M, Caillon-Caravanier M, Dridi Y, Galiano H, Lemordant D (2008) *J Phys Chem B* 112:13335–13343

Nanotoxicology in Green Nanoscience

LEAH WEHMAS, ROBERT L. TANGUAY

Department of Environmental and Molecular Toxicology, Environmental Health Sciences Center, Oregon State University, Corvallis, OR, USA

Article Outline

Glossary
 Definition of the Subject
 Introduction
 Nanotoxicology's Role in Green Nanoscience
 The Embryonic Zebrafish Model
 Nanotoxicology and the Zebrafish Model
 Future Directions
 Bibliography

Glossary

Embryonic development The molecular signaling, cell divisions, cell rearrangements, and cell differentiation that lead to tissues, organs, and structures of an organism.

High throughput A method of stream-lining, often through automation, testing procedures to rapidly conduct thousands of experiments.

Nanotechnology As defined by the National Nanotechnology Initiative, “nanotechnology is the understanding and control of matter at the nano-scale, at dimensions between approximately 1 and 100 nm, where unique phenomena enable novel applications.”

Structure activity relationships (SARS) A method of relating how structural and physiochemical properties of a compound influence biological activity.

Tiered approach An approach that optimizes identification of potentially hazardous compounds through testing and systematic interpretation of results [1].

Toxicology testing Testing to examine and understand the adverse effects of physical, biological, or chemical compounds on organisms and the environment with the objective of mitigation or prevention [2].

Definition of the Subject

Nanotechnology holds great promise for future economical and technological advances, yet health and safety concerns regarding nanomaterials persist. As an emerging technology, nanotechnology is in the unique position to proactively address health and safety concerns throughout the product life cycle. Green chemistry aims to create benign compounds in a way that prevents pollution and reduces waste throughout every stage of production. Through green nanoscience, the principles of green chemistry can be applied toward making high performance, yet inherently safe nanomaterials. Successful application of green chemistry principles to assess nanomaterial health and safety requires efficient, predictive, high-throughput nanotoxicity testing. With these approaches, designers and manufacturers of nanomaterials can assess nanotoxicity early in production to redesign or replace hazardous nanomaterials.

Introduction

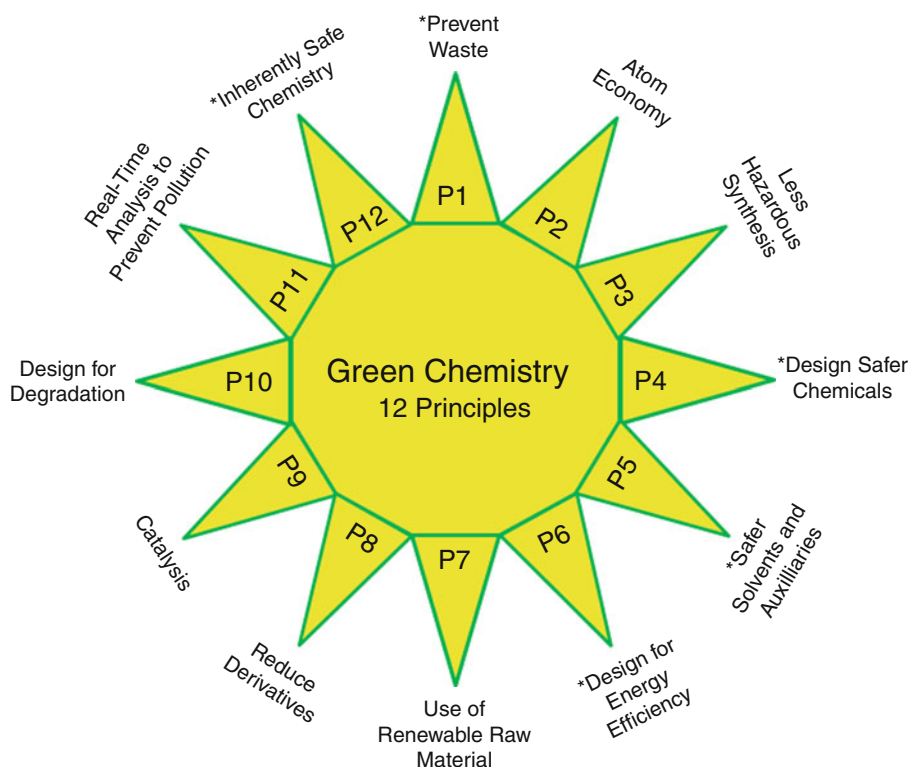
Nanotechnology is a field of technology, engineering, and science that involves shaping and rearranging matter near the atomic level to make materials, particles,

or structures with at least one dimension on the scale of 1–100 nm [3]. At a size that is approximately one billionth that of a meter, distinctive mechanical, optical, and physiochemical properties emerge [3]. Nanotechnology exploits these novel nanoscale properties for unique applications in electronics, targeted drug delivery, personal care products, textiles, etc. [4–9].

The numerous applications and continued innovation in nanotechnology have contributed to the exponential growth in the industry. Within 5 years nanotechnology is projected to expand from a \$3.8 billion industry to a \$2.5 trillion industry by 2015 [10] making it a significant contributor to the global economy. This rapid growth indicates that nanotechnology is transitioning from a discovery to production industry [11, 12]. This scale-up in nanomaterial production, while contributing significantly to economic growth and industrial improvement, will also significantly increase the potential for environmental and human exposure to nanomaterials [13]. Therefore, the swift assessment of nano-environmental and

human safety (nanoEHS) risks is necessary. While numerous nanomaterial toxicity studies have been conducted, conflicting results make it difficult to adequately assess nanoEHS. The unique properties that make nanomaterials an attractive technology (surface chemistry, surface area, size, shape, core material functionalization, aggregation, etc.) may also contribute to novel biological effects as a result of nanomaterial exposure [12]. Toxicology will play an important role in elucidating the mechanisms of those interactions [14] and will be referred to as nanotoxicology when applied to nanomaterials.

The emergence of nanotechnology occurred in tandem with the introduction of the concepts of green chemistry [15] making it a technology ideally suited for incorporation of the principles of green chemistry (Fig. 1) early in the industry [16]. When nanomaterials are considered for commercialization, collaborative work between the producers of nanomaterials and toxicologists is necessary to prioritize [17] and design comprehensive, efficient, high-throughput nanotoxicity



Nanotoxicology in Green Nanoscience. Figure 1

The 12 principles of green chemistry. Asterisks designate principles that can be applied to nanotoxicology

tests to address nanoEHS. Faster and predictive nanotoxicity tests will require an *in vivo* approach at several levels of biological organization including molecular → cellular → tissue → organ → whole organism effects. Toxicity testing implemented early in the nanomaterial design process will allow for refinement or replacement of unsafe nanomaterials earlier in the development pipeline. This will enable the production of inherently safer nanomaterials through the incorporation of the 12 principles of green chemistry [15], a major part of the green nanoscience research approach [16].

This entry will explore the role of toxicology in implementing green nanoscience, the methodology of incorporating nanotoxicology in order to directly and indirectly address the principles of green chemistry in green nanoscience, and the importance of utilizing robust models for nanotoxicity testing.

Nanotoxicology's Role in Green Nanoscience

Application of Green Chemistry

When one thinks of nanotechnology and nanomaterials consumer products such as sunscreens, cosmetics, or antimicrobial socks come to mind, yet there are several innovative nanomaterials that may have environmental remediation applications in oil spill cleanup, water purification, and air filtration [18–20]. As many nanomaterials move closer to commercialization and production levels expand, the potential environmental health applications and economic growth in this field could greatly benefit the society. However, the full benefit and promise of nanotechnology may be greatly impeded, or never realized, until it gains acceptance as safe by the general public and lawmakers. To gain widespread safety acceptance, nanoEHS concerns must be addressed. This will require green nanoscience approaches. Green nanoscience [16] aims to produce inherently safer nanomaterials by combining the 12 principles of green chemistry with nanoscience to provide guidelines to design effective materials with minimized waste and hazard throughout the life cycle [15]. Nanotechnology is a new and evolving industry, and is in the unique position to proactively apply green chemistry principles early in the design and synthesis of nanomaterials.

Material scientists and chemists will not be able to apply all the principles of green chemistry to nanoscience efficiently without collaborations with other scientific disciplines. According to Anastas and Eghbali, there has been a general lack of concern toward addressing principles 4 (P4) of green chemistry which involves designing compounds, such as nanomaterials, in a way to minimize overall hazard [21]. Green nanoscience can avoid the folly of other adopters of green chemistry and alleviate nanoEHS concerns through incorporation of nanotoxicology. The old paradigm of conducting toxicity testing at the end of a development process will be neither sustainable nor suitable to meet the goals of green nanotoxicology. Instead, nanotoxicology must be implemented early and often in green nanomaterial design and synthesis. Nanotoxicology will provide data that can be used in the refinement or replacement of unsafe nanomaterials through direct incorporation of P4 of green chemistry. In addition, it can be used to evaluate the safety of feed stock materials required for the synthesis of nanomaterials such as solvents or other synthesis by-products (P5). This will aid green nanoscience in designing benign nanomaterials (P12). Using nanotoxicology to address P4, P5, and P12 in green nanoscience may be fairly obvious (Fig. 1). What may not be so obvious is how nanotoxicology can be applied to P1 and P6 (Fig. 1) depending on the testing platform and models used which will be covered later in this entry.

Toxicology Models

Traditionally, mammalian toxicity tests are favored and thought to be the most predictive for assessing the human health and safety of compounds. Mammalian models share a high degree of genetic similarity with humans making resulting toxicity data easier to extrapolate to humans [14], thereby addressing P4 of green chemistry. However, concerns about animal welfare have generated a need to restrict mammalian use to only essential testing and to find alternative testing approaches [22]. Maintaining, housing, and testing mammals require significant infrastructure, materials, and time. Cumulatively, these requirements make these tests low throughput, expensive, and impractical for the broad-based nanoEHS research needs. Additionally,

a large amount of test compound is necessary for mammalian toxicity assays making testing nanomaterials particularly challenging as they are often synthesized in small batches during the discovery phase [11]. Larger synthetic needs also create more waste which conflicts with green chemistry P1. Therefore, in vitro assays have become a popular alternative to whole animal tests as infrastructure requirements and maintenance needs are greatly reduced. These assays are also high-throughput and utilize far less test material. While in vitro tests can provide valuable data on a specific pathway or processes, the hazard and risk predictability of these assays for whole animals is quite limited. In vitro assays are inherently less complex as they most often consist of single cell types in cultured dishes. In addition, specific cell types express unique gene products. Since the critical gene products that play roles in nanotoxicological responses are not known in advance, it is a challenge to select the most appropriate in vitro system for nanoEHS research. To increase the sensitivity of in vitro toxicity screens, multiple in vitro assays are assembled to account for routes of exposure, organ effects, chemical deposition, etc. [23]. This reduces the economy of the in vitro approach to toxicity testing as it becomes necessary to conceptually rebuild an organism through numerous assays to obtain the required complexity for predictive results. These caveats will make it challenging to obtain reliable nanoEHS results in the near future. As eluded to in the introduction, the challenges of green nanoscience in conducting nanotoxicity testing for rapid, comprehensive assessment of nanoEHS necessitates the use of a robust and predictive in vivo model that can be adapted to high-throughput screening. Here the embryonic zebrafish (*Danio rerio*) is proposed as the ideal model for nanotoxicity tests.

The Embryonic Zebrafish Model

Background

There are several basic attributes of the embryonic zebrafish model that make it ideal for nanotoxicology and green nanoscience. The husbandry requirements of the zebrafish are minimal. As a small vertebrate model with high fecundity (200–300 eggs per day), short intervals between female spawns (2–3 days), and quick maturation, many zebrafish can be housed in

a small area reducing infrastructure costs [24]. Thousands of eggs can be produced daily for experimentation. Aside from ease of care and maintenance, additional benefits of the zebrafish model include transparent eggs and external embryogenesis which facilitates examination of early developmental processes in real time [24].

The transparency of embryos during development make the zebrafish model exceptionally well suited to high-throughput testing platforms which will be necessary to alleviate backlog in demand for quick nanotoxicity evaluations. During early development complex molecular signaling events and organogenesis make the embryo highly susceptible to environmental perturbations [14]. Exposure to nanomaterials at this critical life stage can disrupt organ function, cellular pathfinding, and/or tissue morphology which can be easily tracked in the transparent zebrafish embryos throughout the duration of a 5-day exposure, noninvasively [25].

Molecular Applications What began as a vertebrate model of developmental genetics is now extensively used a model for infectious diseases and immune system function [26–28], to understand hematopoiesis and related diseases [28, 29], for aging [30, 31], oncology [32–34], cardiovascular development [35, 36], kidney development and disease [37–39], for identifying the etiology of diseases of the eye [40, 41] and ear [42, 43], for understanding mechanisms of drug abuse [44–46], and for regenerative medicine [47–49]. Importantly, the zebrafish genome has been sequenced [50] and this information along with the conserved anatomy and physiology between fish and humans is being exploited by the biomedical research community (reviewed in [51–55]). Clearly, there are exciting possibilities to using zebrafish as a model for mechanistic-based nanotoxicity studies.

With the remarkable conservation between the zebrafish and human genome, it is anticipated that nearly all the toxicologically relevant pathways also function in zebrafish. The most well-studied toxicologically relevant pathway in zebrafish is the aryl hydrocarbon receptor (AHR) signal transduction pathway. Over the past few years, the zebrafish AHR pathway has been identified and completely characterized (reviewed in [56–58]). This has served as proof of concept and

demonstrates that toxicology pathways can be efficiently examined in zebrafish.

A first step in elucidating the role of a gene in a biological process or in mediating responses to external stimuli (i.e., toxic responses) is to manipulate the expression level of the gene product. Producing knockout mice using embryonic stem cells has revolutionized biomedical research [59] and has been useful for toxicological studies in rodents [60–62]. Despite extensive efforts, embryonic stem cells for the zebrafish have yet to be developed [63], but a number of approaches have been effectively used to define the role of individual genes in zebrafish. The most common approach is the use of morpholinos which are chemically modified oligonucleotides that are resistant to chemical and enzymatic degradation yet maintain base pairing properties [64]. Morpholinos have been shown to bind to and block translation of mRNA in embryonic zebrafish [65–67]. Morpholinos can also be designed to bind to complement sequences at intron-exon splice junctions to inhibit proper mRNA processing. This approach is especially effective in zebrafish embryos [68–71]. Splice blocking morpholinos are useful to target genes that have alternative translation start sites, and to generate misspliced transcripts that result in dominant-negative proteins [68]. The successful application of morpholinos in zebrafish is evident by the large number of published studies using this approach (<http://www.gene-tools.com/Publications>).

With the availability of the entire zebrafish genome, conceivably any zebrafish gene can be targeted individually, or in combinations, using morpholinos and other antisense techniques. Thus, the function of mammalian orthologs can be rapidly evaluated in zebrafish. There are, however, limitations to these approaches. The gene repression effects are transient, which are most suitable for evaluating gene functions during development, (up to approximately 3–5 days post fertilization). Efforts to effectively deliver morpholinos to zebrafish at older life stages or to specific cellular targets are not routine. To overcome this limitation, heritable knockout approaches have been developed.

Despite the lack of a technology for gene knockout by homologous recombination, researchers have generated a large number of mutant fish lines. The majority of these were created using the chemical mutagen, ethylnitrosourea (ENU), followed by

phenotypic screening for developmental defects [72, 73]. Such screens facilitate the identification of genes essential for biological processes with no prior functional knowledge of the genes. To date, large-scale zebrafish mutant searches have led to the identification of thousands of mutants with unique embryonic and larval development defects [74, 75]. Toxicologists can use available mutants that phenocopy the effects of toxicant exposure to help identify molecular targets. Furthermore, since genome-wide genetic screens are routine in zebrafish, there is a significant untapped potential to use forward genetics to identify modifiers of susceptibility to xenobiotic insult. Since chemical mutagenesis is non-biased, there is a high probability that novel targets can be identified in zebrafish. For instance, a genetic screen for increased susceptibility to nanomaterial exposure could be used to identify suppressor pathways. Screening for resistance to chemical exposure has the potential to identify genes that modulate the toxic response. With the power of comparative genomics, the identification of novel genes can be rapidly evaluated using integrated approaches in mammals. An advantage of ENU-mediated mutagenesis is that small mutations are created; however, the identification of the mutated gene remains a challenge. The once tedious and expensive approach of positional cloning has successfully been used to identify hundreds of the mutated genes [76, 77], and has evolved into a more efficient “targeted” methodology. For example, the Targeting Induced Local Lesions in Genomes (TILLING) method can now be used in zebrafish to identify and/or select for loss-of-function alleles (knockouts) [78]. TILLING is conceptually simple; random mutations are created using traditional chemical mutagenesis followed by high-throughput PCR-based screening to identify point mutations in the gene of interest. Therefore, TILLING is a targeted approach and, in theory, every zebrafish gene can now be knocked out. Insertional mutagenesis is an alternative approach to generate zebrafish mutants and has the benefit that the interrupted gene can be immediately identified [79–81]. An emerging method utilizing zinc finger nucleases (ZFNs) holds promise in bypassing the inability to generate knockout zebrafish by homologous recombination. Zinc finger nucleases are engineered to bind and cleave specific sections of DNA. Application of ZFNs has been successful in

generating targeted germline mutations in the zebrafish [82].

Numerous techniques have been developed to introduce foreign DNA into zebrafish including micro-injection [83], retroviral infection [79, 80], biolistic delivery [84], and electroporation [85]. The applications for transgenic zebrafish to unravel mechanisms of toxicity appear limitless. Conditional promoters such as the hsp70 [86] and tetracycline responsive [87] can be used to control transgene expression. Functional analysis in zebrafish can also be achieved using ectopic expression vectors under the control of tissue-specific promoters reviewed in Refs. [88, 89]. The number of transgenic zebrafish lines is rapidly growing with numerous lines useful for toxicological research. For example, transgenic fish that express fluorescent reporter genes in response to stimuli have been used to evaluate estrogenic chemical exposure [90], to visualize calcium signaling in vivo [91, 92], to respond to AHR ligands [93], etc. In addition to assessing the transcriptional responses to xenobiotics, fluorescent transgenes have proven valuable to monitor the dynamic movements of cells. Since embryonic zebrafish are transparent, toxicity targets can be non-invasively monitored, in real time. Examples include the use of transgenic lines with neuronal-specific expression of GFP [94] to assess the impact of nicotine on motor neuron development [45], the use of sonic hedgehog responsive fish to evaluate the impact of TCDD on CNS development [95], and the use of heat shock responsive fish to localize the targets of cadmium toxicity [96]. Finally, transgenic zebrafish lines have been developed to detect mutations in vivo [97].

An emerging area is the use of zebrafish as a model to elucidate the impact of xenobiotics on CNS development and function. Zebrafish possess all typical senses including vision, olfaction, taste, touch, balance, and hearing; and their sensory pathways share an overall homology with humans [98]. Cognitive behavioral tests suggest that anatomic substrates of cognitive behavior are also conserved between fish and other vertebrates. Thus, similar to observations of hippocampal lesions in mammals, lesions of the structural homolog of the hippocampus in fish selectively impair spatial memory [99]. Zebrafish behavior is altered by exposure to ethanol [46, 100, 101], nicotine [45], cocaine [44], antipsychotics [102], pesticides [103, 104], and

metals such as mercury [105]. Recent studies indicate that broad-based chemical genetic screens can be used to identify novel, clinically relevant neuroactive compounds [106]. These initial behavioral studies suggest that the zebrafish model may be ideal to investigate the long-term consequences on CNS function following transient exposures to xenobiotics.

Similar to other toxicological models, “omic” based approaches are now routine using zebrafish. With the availability of the genome sequence, custom and commercial cDNA and oligonucleotide microarrays are currently available. These microarrays have been effectively used to define developmental stage and tissue-specific transcriptomes [107–112] and to identify gene expression changes in response to xenobiotic exposure. Microarray technology has been successfully applied to toxicology. For instance, the gene expression changes in response to TCDD [49, 58, 113], hypoxia [114], 4-nonylphenol [115], and chlorpromazine [110] have been reported. Further understanding of genome-wide changes in transcript levels will be possible through application of RNA sequencing technology [116] to examine how the entire zebrafish transcriptome changes with exposure to toxicants. The amenability of the zebrafish to all the molecular biology techniques recently described will be very important in determining the mechanisms by which nanomaterials can cause adverse reactions in humans and other organisms.

High-Throughput Platform The extensive growth in the field of nanotechnology has created a pipeline problem for efficient nanoEHS assessment. High-throughput nanotoxicity testing approaches can alleviate this problem. The embryonic zebrafish model is already a moderate-throughput platform, but readily amenable to high-throughput nanotoxicity testing with the incorporation of robotics and automation.

Zebrafish embryos develop in a chorion, which is an acellular envelope that acts as a protective barrier. Normally, when reared at 28°C, zebrafish hatch at between 48 and 72 hour post fertilization (hpf). To increase the sensitivity of the embryonic assay, it is critical to initiate exposures as early as possible. The earliest feasible stage to remove the chorions from the zebrafish embryos is at 6 hpf via enzymatic digestion. The methods for removing the chorion from large

groups of embryos have been optimized, eliminating a potential barrier to uptake from the surrounding water [25].

A potential bottleneck to high-throughput embryonic zebrafish assays is the placement of dechorionated zebrafish embryos into 96- or 384-well assay plates. For toxicity exposures, it is necessary to sort embryos by identifying and selecting healthy eggs at similar developmental stage. This is a delicate process because the embryos can easily rupture. Incorporation of precision robotic plate loaders with visual systems to sort embryos could greatly increase the productivity of the embryonic zebrafish assay. The development of such a system has recently been achieved [117]. Microinjection of embryos is a molecular technique that introduces exogenous chemicals, nanomaterials, DNA, RNA, protein, or synthetic oligonucleotides into zebrafish embryos. These techniques are powerful investigative tools that are time intensive. Precision robotics with computer vision and control could hasten this arduous task and increase injection success rates [118].

Perhaps the greatest hurdle in developing high-throughput embryonic zebrafish nanotoxicity tests involves screening for sometimes subtle phenotypic changes. Zebrafish embryos can display a wide range of malformations that must be evaluated by laboratory personnel [25]. Experienced toxicologists may require up to 30 min per plate depending on the number of endpoints investigated. Introduction of automated, image-based screening methods would greatly enhance throughput. After devising automated plate preparation and embryo handling, the next step to developing a high-throughput assay is to set up imaging systems that can automatically focus and capture images [119]. There must be sufficient data storage capabilities to save thousands of high content images. In addition, software algorithms are necessary to process the image data [119], and the software must be capable of efficiently evaluating phenotypic changes. Commercial software analyzing *in vitro* cell cultures is currently available [119]; however, the embryonic zebrafish presents an additional challenge of orientation of the embryo which can influence image content. Finally, data mining and statistical analyses tools are needed to distinguish and report significant results which can then be confirmed and investigated further if

necessary [119]. This will require a high amount of computing power. At this point in automation, researchers must pay special attention to the quality of the data analysis such as minimizing the generation of false positives and negatives [120]. Addressing the steps necessary to high-throughput assay development will enhance the efficiency of the zebrafish assay while resulting in more objective data. This will help address the nanoEHS concerns [119]. Currently, there are laboratories that are independently developing such automated systems.

The Zebrafish as a Green Model

Since the development of the 12 principles of green chemistry in the early 1990s [15], the movement toward adopting these principles in research has increased [21]. Many of the aforementioned benefits of the embryonic zebrafish model make it ideal for use in green toxicology and nanotoxicology. This section will discuss how the zebrafish model applies to the principles of green chemistry.

While the principles of green chemistry generally refer to synthesis and design of chemical compounds, they can be adapted to non-chemistry applications. Of the 12 principles of green chemistry (Fig. 1), use of the zebrafish model can help green nanoscience address P1 and P6. Principle 1 of green chemistry pertains to reduction in waste (Fig. 1) and use of the zebrafish as an *in vivo*, vertebrate model reduces the space, use of materials and equipment, and generation of waste required for maintaining mammals. For example, in our laboratory a standard nanotoxicity assay involves performing a concentration-response exposure in three 96-well plates with 100 μ L of solution per well. Each plate contains an $n = 16$ per treatment and controls that contain no test material. The chemical treatments include a fivefold dilution series; therefore, an exposure with three plates of five treatments uses low milligram quantities of exposure material reducing waste generation (P1). While the zebrafish cannot completely replace the use of mammalian testing, it can be used as a high-throughput screen to prioritize nanomaterials for mammalian tests which will reduce their use.

With the great demand for toxicity testing on many new and existing compounds, mammalian testing cannot keep pace. Therefore, introduction of the zebrafish

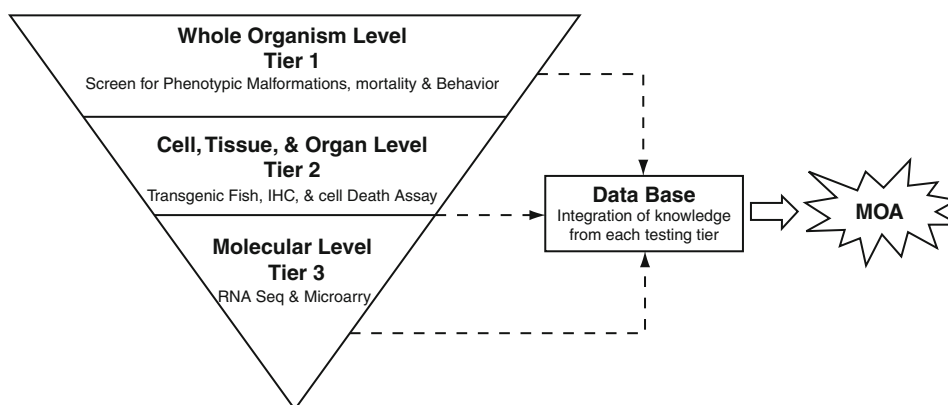
assay can increase the energy efficiency of toxicity testing in accord with P6 of green chemistry (Fig. 1). Maintaining and conducting toxicity tests on mammals requires much more energy input than zebrafish toxicity testing. Even with *in vitro* systems, energy is required in maintaining the cells and conducting the tests. If numerous *in vitro* tests are required to obtain human-relevant toxicity data, then the energy efficiency of zebrafish testing is superior.

In one assay, the zebrafish model effectively combines the predictability of mammalian toxicity tests with the speed and economy of *in vitro* assays. Zebrafish embryonic assays are increasingly becoming high-throughput through automation and development. As the molecular signaling cascades, physiology, and anatomy are well conserved between humans and zebrafish, nanotoxicity testing in zebrafish should be highly relevant. To date, many phenotypic endpoints of toxicity in zebrafish can be applied to understanding conditions in humans. For instance, the induction of bradycardia in zebrafish embryos has been applied to identify pre-clinical drugs with risks for heart arrhythmias in humans [121]. Berghmans et al. demonstrated a correlation between the effects of pharmacologically active compounds in humans or mammals with the effects in embryonic zebrafish through cardiac, intestinal, and visual assays [122]. These results support the

zebrafish model's predictive capabilities and utility to help green nanoscience achieve the aim of designing inherently safer nanomaterials (P12).

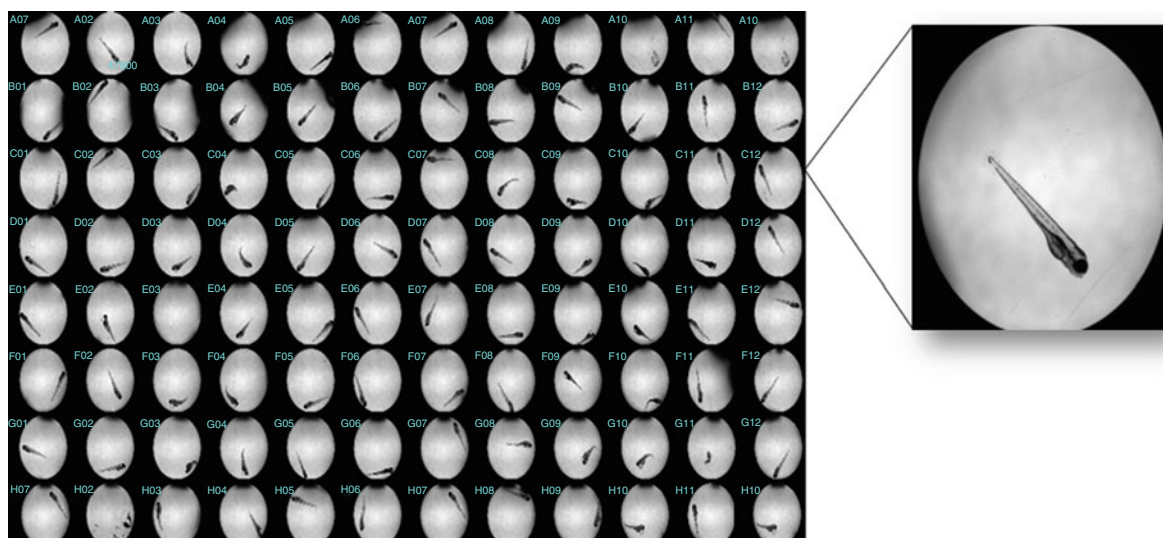
Nanotoxicology and the Zebrafish Model

The zebrafish model has been used to design a high-throughput, three-tiered testing approach that evaluates adverse effects at several levels of organization (Fig. 2). While this tiered testing scheme evaluates the safety of nanomaterials, the major objective of this assay is to understand the physiochemical properties of nanomaterials that result in toxicity. A critical first step is the availability of well-characterized nanomaterials, where the size, number of atoms, agglomeration state, shape, surface chemistry, charge, and purity are precisely known [11, 123]. Tier one involves a waterborne exposure of dechorionated zebrafish embryos to the well-characterized nanomaterials to evaluate effects at the whole organism level. The exposure is continuous for 5 days beginning at 6 hpf and finishing at 120 hpf. During the assay, zebrafish embryos are screened for phenotypic effects at two time points (24 and 120 hpf) which manifest as mortality or phenotypic changes including increased spontaneous movement, developmental delays, or notochord malformations at 24 hpf. At 120 hpf, viable larval fish are screened for pericardial edema; yolk sac



Nanotoxicology in Green Nanoscience. Figure 2

Diagram outlining the tiers of nanotoxicity testing (1–3) and the types the laboratory techniques used to understand effects at the different levels of organization. Data from all three tiers are integrated to determine the mode of action (MOA) of a nanomaterial



Nanotoxicology in Green Nanoscience. Figure 3

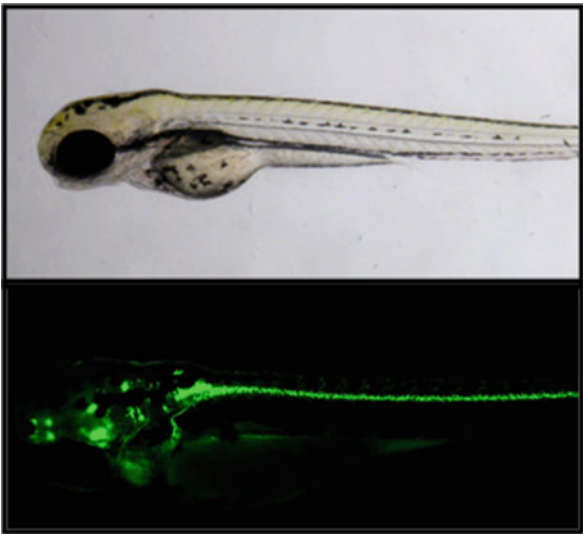
Image of a 96-well plate which is used in the three tiers of zebrafish nanotoxicity assay. The plate contains zebrafish larva (120 hpf) after 5 days of exposure. *Right image* shows an enlargement of a single zebrafish in an exposure well

edema; or eye, nose, snout, jaw, axis, fin, brain, otic, notochord, or circulatory malformations using multi-well microscope imaging (Fig. 3). To evaluate impacts on the CNS development and behavior, two assays have been developed for screening at 120 hpf. The effects on motility and tactile response are evaluated by determining if fish display normal predator avoidance behavior by swimming away from a light touch to the head or tail. The second behavioral assay exploits the well-studied phenomenon of increased larval zebrafish motor activity in the dark [124]. The locomotor activity of nanomaterial exposed zebrafish larvae is recorded using Videotrack V3 software (Viewpoint Life Sciences, Inc., France) for 10 min in the dark following a 10 min rest period in the light. The results reveal whether nanomaterial exposure impacts this complex light-induced behavior.

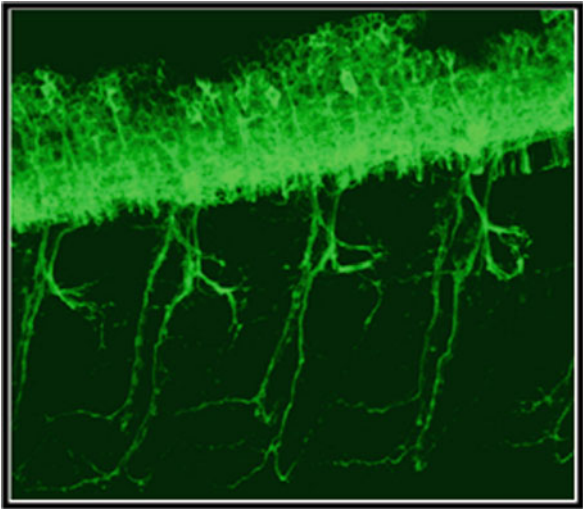
To understand zebrafish nanomaterial absorption, it is necessary to experimentally measure nanomaterial uptake so that the relationship between in vivo nanoparticle dosimetry and response can be modeled. Determination of the internal dose of nanoparticles in embryonic zebrafish requires the use of Instrumental Neutron Activation Analysis (INAA) or Inductively Coupled Plasma-Optical Emission Spectrometry/Mass Spectrometry (ICP-OES/ICP-MS respectively).

Nanomaterials that elicit phenotypic or behavioral responses in the zebrafish advance to tier two of testing to elucidate effects occurring at the cellular, tissue, or organ level. Use of transgenic zebrafish with fluorescent reporters and/or immunohistochemistry can help pinpoint the transport and deposition of a nanomaterial by tracking where adverse effects occur. Examples of two transgenic lines are illustrated in Figs. 4 and 5. Figure 4 illustrates a live 4-day-old zebrafish under non-excitatory light (top panel), and the same live fish images under excitatory light (bottom panel). The green fluorescent protein specifically labels the secondary motor neurons in these fish. Figure 5 illustrates a high magnification view of the spinal cord of a live zebrafish illustrating the detailed structures of the spinal neurons as they innervate into the adjacent muscle. These and other transgenic fish lines are useful tools when nanomaterial exposures are suspected to disrupt either the development or migration of spinal neurons. Another in vivo tier-2 assay is to determine if nanomaterial exposure results in increased cell death (Fig. 6). In this example, exposure to fullerenes leads to an increase in cellular death throughout the live zebrafish embryo [125].

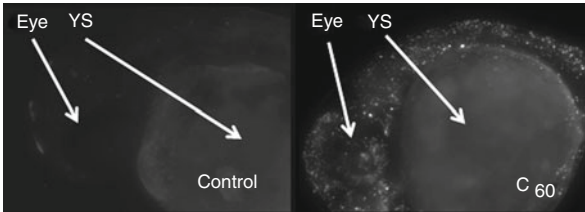
Tier three testing focuses on the molecular effects of nanomaterial exposure. RNA seq or microarrays can be



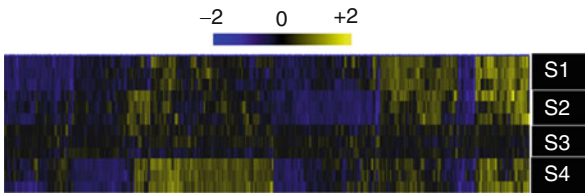
Nanotoxicology in Green Nanoscience. Figure 4 Live 96-hour-old transgenic zebrafish (islet 1-gfp) that expresses the green fluorescent protein expressed specifically in spinal secondary motor neurons (*bottom image*). *Top image* is the transgenic zebrafish imaged under non-excitatory light



Nanotoxicology in Green Nanoscience. Figure 5 High-resolution confocal image of live transgenic zebrafish (nbt-gfp) expressing the green fluorescent protein in primary motor neurons



Nanotoxicology in Green Nanoscience. Figure 6 Representative example of live 24-hour-old zebrafish embryos stained with acridine orange (AO). AO stains cells undergoing cellular death assay. AO binds to nucleic acids and fluoresces orange when cells enter lysosomes for degradation. Greater fluorescence in the nanoparticle (carbon 60 fullerene-C₆₀) exposed embryo indicates enhanced cellular death compared to the control. YS yolk sac



Nanotoxicology in Green Nanoscience. Figure 7 Representative microarray heat map displaying changes in gene expression following exposure to four different chemicals. The rows across the heat map (labeled S1, S2, S3, and S4) represent three different samples (in triplicate), while each column indicates a different gene. Blue signifies a decrease in gene expression, black signifies no changes, and yellow signifies an increase in gene expression

used to evaluate changes in gene expression profiles. Figure 7 illustrates the power and complexity of this approach. Following exposure to four different chemicals, the impact of the exposures on all genes in the genome was assessed, and provided distinct changes in gene expression. Known gene relationships (ontology) can be used to find particular pathways that may be impacted by nanomaterial exposure. The information from tiers two and three also enter the database. Data across all three tiers of nanotoxicity testing are integrated to define the mode of nanomaterial toxicity.

Future Directions

Regardless of the many marketable nanomaterials and nanomaterial-based products poised for industrial production, few have actually been commercialized. A partial explanation for this lag can be assigned to difficulty in scaling up the nanosynthesis process from bench level to production level and still maintaining the quality, purity, and efficiency of the small batches [12]. However, a lack of nanoEHS data for confident assessment of the risks posed by exposure to nanomaterials remains a significant problem. The most critical need is to identify the design rules that define specific physiochemical properties of nanomaterials that predict biologic effects (biotic-nano models based on structure activity relationships) [126].

A major bottleneck to ultra high-throughput screening in the zebrafish assay will be caused by time constraints in data processing resulting from the large volume of data generated from high content imaging and large sample sizes. To avoid increases in total handling time of the screening process, increased computing power and reduction in the time required for data analyses and interpretation of the results are necessary [120]. Improvements in data analyses and interpretation will necessitate improvements and innovations in computer software. Improvements in computer software and whole organism-based imaging technology will also be necessary in augmenting the sensitivity of the zebrafish assay. For example, improvements in and increased dissemination of Cognition Network Technology (CNT) image analysis, a method that attempts to imitate human thought processes, over pixel-based image analysis methods could potentially allow for capturing subtle phenotypes that cannot be seen by the human eye irrespective of the orientation of the embryo [19]. Another method of increasing the sensitivity of the zebrafish assay may be through the addition of a high-throughput adaptation of micro-CT imaging. Micro-CT imaging, which is already used in mammalian research, is able to capture three-dimensional images of an organism noninvasively [127]. This technology could be used to capture slight but perhaps significant morphological responses that are presently not practical to assess on large numbers of animals.

Analysis of the changes in gene expression that occur with toxicant exposure has become a powerful

tool in toxicology for determining toxicity mechanisms. With the development of RNA-Seq and organism-wide changes in gene transcripts will allow faster identification of molecular targets of compounds [116]. RNA-Seq also generates large amounts of data which must be processed and analyzed. Future work must be directed at determining the best way to analyze RNA-Seq data and in automating the processing and analysis procedures. This will assist in timely gene expression analysis of effects resulting from nanomaterial exposure in the zebrafish assay.

Adoption of high-throughput embryonic zebrafish assay will alleviate the backlog of nanotoxicity tests necessary for nanomaterial risk assessment, which will not be feasible to evaluate the toxicity of all nanomaterials [128]. Therefore, it is imperative that nanotoxicity screens be conducted in a way to accurately link the physiochemical properties of nanomaterials to toxicological endpoints, effectively creating structure activity relationships (SARS) [126]. By developing nano-SARS, the toxicity of new nanomaterials becomes increasingly predictive of human safety. It is realistic to say that extensive development of nano-SARS will occur over the next decade. There are collaborative efforts ongoing for establishing the foundation for such models [126]. For example, a few nanotoxicity studies have demonstrated that surface chemistry [129, 130] and size [129, 131, 132] may be the most important predictors of nanomaterial activity.

A framework must be also created to organize and curate the data for nano-SARS. To be successful, the framework must include three key guidelines. First, the nanomaterials tested in the toxicity assays must be well characterized [11, 123]. Accurate characterization of nanomaterials remains a major immediate objective in green nanoscience. Second, libraries of well-characterized nanomaterials with systematic variations in structural or physiochemical properties must be assembled. By screening these libraries in the embryonic zebrafish assay and focusing on one property at a time, one can begin to elucidate how nanomaterial core size, core composition, and functionalization influence biological activity [128]. Third, the data gathered from these tests must be assembled into an open database where others can access the information. The success of this framework requires close collaborative

efforts between toxicologists, nanomaterial scientists, regulatory agencies, and industry. With the use of efficient vertebrate toxicological models like zebrafish and the development of accurate SARS, benign by design nanomaterials will be closer to realization.

Bibliography

- Hushon J, Clerman R, Wagner B (1979) Tiered testing for chemical hazard assessment. *Environ Sci Technol* 13:1202–1207
- Society of Toxicology (2005) How do you define toxicology? *Soc Toxicol Commun.* http://www.toxicology.org/ai/pub/si05/SI05_Define.asp
- NSET/NEHI (2011) NNI environmental, health, and safety research strategy fact sheet. http://nano.gov/sites/default/files/pub_resource/2011_ehs_strategy_fact_sheet_locked.pdf
- Forrest DR (2001) Molecular nanotechnology. *IEEE Instrum Meas Mag* 4(3):11–20
- Lecoanet H, Wiesner MR (2004) Assessment of the mobility of nanomaterials in groundwater aquifers. *Abs Pap Am Chem Soc* 227:U1275–U1275
- Lecoanet HF, Bottero JY, Wiesner MR (2004) Laboratory assessment of the mobility of nanomaterials in porous media. *Environ Sci Technol* 38:5164–5169
- Lecoanet HF, Wiesner MR (2004) Velocity effects on fullerene and oxide nanoparticle deposition in porous media. *Environ Sci Technol* 38:4377–4382
- Okamoto Y (2001) Ab initio investigation of hydrogenation of C-60. *J Phys Chem A* 105:7634–7637
- Sun O, Wang Q, Jena P, Kawazoe Y (2005) Clustering of Ti on a C-60 surface and its effect on hydrogen storage. *J Am Chem Soc* 127:14582–14583
- Lux Research (2009) The recession's ripple effect on nanotech. Lux Research Inc., New York
- Dahl J, Maddux BLS, Hutchison JE (2007) Green nanosynthesis. *Chem Rev* 107:2228–2269
- Hutchison JE (2008) Greener nanoscience: a proactive approach to advancing applications and reducing implications of nanotechnology. *ACS Nano* 2:395–402
- Thomas K, Sayre P (2005) Research strategies for safety evaluation of nanomaterials, Part I: evaluating the human health implications of exposure to nanoscale materials. *Toxicol Sci* 87:316–321
- NRC (2000) Scientific Frontiers in developmental toxicology and risk assessment. National Academy Press, Washington, DC, pp 1–327
- Anastas PT, Warner JC (1998) Green chemistry: theory and practice. Oxford University Press, Oxford/New York, xi, 135 p
- McKenzie LC, Hutchinson J (2004) Green nanoscience: an integrated approach to greener products, processes and applications. *Chem Today* 22:30–33
- Abbott BD, Perdew GH, Buckalew AR, Birnbaum LS (1994) Interactive regulation of Ah and glucocorticoid receptors in the synergistic induction of cleft palate by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin and hydrocortisone. *Toxicol Appl Pharmacol* 128:138–150
- Podgórski A, Balazy A, Gradon L (2006) Application of nanofibers to improve the filtration efficiency of the most penetrating aerosol particles in fibrous filters. *Chem Eng Sci* 61:6804–6815
- Savage N, Diallo MS (2005) Nanomaterials and water purification: opportunities and challenges. *J Nanoparticle Res* 7: 331–342
- Yuan J, Liu X, Akbulut O, Hu J, Suib SL et al (2008) Superwetting nanowire membranes for selective absorption. *Nat Nano* 3:332–336
- Anastas P, Eghbali N (2010) Green chemistry: principles and practice. *Chem Soc Rev* 39:301–312
- Lein P, Silbergeld E, Locke P, Goldberg AM (2005) In vitro and other alternative approaches to developmental neurotoxicity testing (DNT). *Environ Toxicol Pharmacol* 19:735–744
- Oberdorster G, Maynard A, Donaldson K, Castranova V, Fitzpatrick J et al (2005) Principles for characterizing the potential human health effects from exposure to nanomaterials: elements of a screening strategy. *Part Fibre Toxicol* 2:8
- Detrich HW, Westerfield M, Zon LI (eds) (1999) The zebrafish biology. Academic, San Diego, 391 p
- Truong L, Harper SL, Tanguay RL (2011) Evaluation of embryotoxicity using the zebrafish model. *Methods Mol Biol* 691:271–279
- van der Sar AM, Appelmelk BJ, Vandenbroucke-Grauls CM, Bitter W (2004) A star with stripes: zebrafish as an infection model. *Trends Microbiol* 12:451–457
- Trede NS, Langenau DM, Traver D, Look AT, Zon LI (2004) The use of zebrafish to understand immunity. *Immunity* 20:367–379
- Traver D, Herbomel P, Patton EE, Murphey RD, Yoder JA et al (2003) The zebrafish as a model organism to study development of the immune system. *Adv Immunol* 81:253–330
- de Jong JL, Zon LI (2005) Use of the zebrafish to study primitive and definitive hematopoiesis. *Annu Rev Genet* 39:481–501
- Gerhard GS (2003) Comparative aspects of zebrafish (*Danio rerio*) as a model for aging research. *Exp Gerontol* 38:1333–1341
- Keller ET, Murtha JM (2004) The use of mature zebrafish (*Danio rerio*) as a model for human aging and disease. *Comp Biochem Physiol C Toxicol Pharmacol* 138:335–341
- Spitsbergen J, Kent M (2003) The state of the art of the zebrafish model for toxicology and toxicologic pathology research – advantages and current limitations. *Toxicol Pathol* 31:62–87
- Amatruda JF, Shepard JL, Stern HM, Zon LI (2002) Zebrafish as a cancer model system. *Cancer Cell* 1:229–231
- Moore JL, Gestl EE, Cheng KC (2004) Mosaic eyes, genomic instability mutants, and cancer susceptibility. *Methods Cell Biol* 76:555–568

35. Chen JN, Fishman MC (2000) Genetic dissection of heart development. *Ernst Schering Res Found Workshop* 29:107–122
36. Beis D, Bartman T, Jin SW, Scott IC, D'Amico LA et al (2005) Genetic and cellular analyses of zebrafish atrioventricular cushion and valve development. *Development* 132: 4193–4204
37. Drummond IA (2004) Zebrafish kidney development. *Methods Cell Biol* 76:501–530
38. Hentschel DM, Park KM, Cilenti L, Zervos AS, Drummond I et al (2005) Acute renal failure in zebrafish: a novel system to study a complex disease. *Am J Physiol Renal Physiol* 288:F923–F929
39. Drummond IA (2005) Kidney development and disease in the zebrafish. *J Am Soc Nephrol* 16:299–304
40. Bahadori R, Huber M, Rinner O, Seeliger MW, Geiger-Rudolph S et al (2003) Retinal function and morphology in two zebrafish models of oculo-renal syndromes. *Eur J Neurosci* 18:1377–1386
41. McMahon C, Semina EV, Link BA (2004) Using zebrafish to study the complex genetics of glaucoma. *Comp Biochem Physiol C Toxicol Pharmacol* 138:343–350
42. Whitfield TT (2002) Zebrafish as a model for hearing and deafness. *J Neurobiol* 53:157–171
43. Nicolson T (2005) The genetics of hearing and balance in zebrafish. *Annu Rev Genet* 39:9–22
44. Darland T, Dowling JE (2001) Behavioral screening for cocaine sensitivity in mutagenized zebrafish. *Proc Natl Acad Sci USA* 98:11691–11696
45. Svoboda KR, Vijayaraghavan S, Tanguay RL (2002) Nicotinic receptors mediate changes in spinal motoneuron development and axonal pathfinding in embryonic zebrafish exposed to nicotine. *J Neurosci* 22:10731–10741
46. Gerlai R, Lahav M, Guo S, Rosenthal A (2000) Drinks like a fish: zebra fish (*Danio rerio*) as a behavior genetic model to study alcohol effects. *Pharmacol Biochem Behav* 67:773–782
47. Poss KD, Keating MT, Nechiporuk A (2003) Tales of regeneration in zebrafish. *Dev Dyn* 226:202–210
48. Akimenko MA, Mari-Beffa M, Becerra J, Geraudie J (2003) Old questions, new tools, and some answers to the mystery of fin regeneration. *Dev Dyn* 226:190–201
49. Andreasen EA, Mathew LK, Tanguay RL (2006) Regenerative growth is impacted by TCDD: gene expression analysis reveals extracellular matrix modulation. *Toxicol Sci* 92:254–269
50. Vogel G (2000) Genomics. Sanger will sequence zebrafish genome. *Science* 290:1671
51. Zon LI (1999) Zebrafish: a new model for human disease. *Genome Res* 9:99–100
52. Ackermann GE, Paw BH (2003) Zebrafish: a genetic model for vertebrate organogenesis and human disorders. *Front Biosci* 8:d1227–d1253
53. Rubinstein AL (2003) Zebrafish: from disease modeling to drug discovery. *Curr Opin Drug Discov Devel* 6:218–223
54. Wixon J (2000) Featured organism: *Danio rerio*, the zebrafish. *Yeast* 17:225–231
55. Dodd A, Curtis PM, Williams LC, Love DR (2000) Zebrafish: bridging the gap between development and disease. *Hum Mol Genet* 9:2443–2449
56. Hahn M (2002) Aryl hydrocarbon receptors: diversity and evolution(1). *Chem Biol Interact* 141:131
57. Tanguay RL, Andreasen EA, Walker MK, Peterson RE (2003) Dioxin toxicity and aryl hydrocarbon receptor signaling in fish. In: Schecter A (ed) *Dioxins and health*. Plenum Press, New York, pp 603–628
58. Carney SA, Chen J, Burns CG, Xiong KM, Peterson RE et al (2006) AHR activation produces heart-specific transcriptional and toxic responses in developing zebrafish. *Mol Pharmacol* 70:549–561
59. Muller U (1999) Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. *Mech Dev* 82:3–21
60. Ryffel B (1997) Impact of knockout mice in toxicology. *Crit Rev Toxicol* 27:135–154
61. Rudolph U, Mohler H (1999) Genetically modified animals in pharmacological research: future trends. *Eur J Pharmacol* 375:327–337
62. Gonzalez FJ (2002) Transgenic models in xenobiotic metabolism and toxicology. *Toxicology* 181–182:237–239
63. Fan L, Collodi P (2002) Progress towards cell-mediated gene transfer in zebrafish. *Brief Funct Genomic Proteomic* 1:131–138
64. Summerton J, Weller D (1997) Morpholino antisense oligomers: design, preparation, and properties. *Antisense Nucleic Acid Drug Dev* 7:187–195
65. Nasevicius A, Ekker SC (2000) Effective targeted gene 'knock-down' in zebrafish. *Nat Genet* 26:216–220
66. Nasevicius A, Ekker SC (2001) The zebrafish as a novel system for functional genomics and therapeutic development applications. *Curr Opin Mol Ther* 3:224–228
67. Nasevicius A, Larson J, Ekker SC (2000) Distinct requirements for zebrafish angiogenesis revealed by a VEGF-A morphant. *Yeast* 17:294–301
68. Draper BW, Morcos PA, Kimmel CB (2001) Inhibition of zebrafish fgf8 pre-mRNA splicing with morpholino oligos: a quantifiable method for gene knockdown. *Genesis* 30:154–156
69. Yan YL, Miller CT, Nissen RM, Singer A, Liu D et al (2002) A zebrafish sox9 gene required for cartilage morphogenesis. *Development* 129:5065–5079
70. Knight RD, Nair S, Nelson SS, Afshar A, Javidan Y et al (2003) Lockjaw encodes a zebrafish tfap2a required for early neural crest development. *Development* 130:5755–5768
71. Imamura S, Kishi S (2005) Molecular cloning and functional characterization of zebrafish ATM. *Int J Biochem Cell Biol* 37:1105–1116
72. Haffter P, Granato M, Brand M, Mullins MC, Hammerschmidt M et al (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* 123:1–36
73. Driever W, Solnica-Krezel L, Schier AF, Neuhauss SC, Malicki J et al (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* 123:37–46

74. Abdelilah S, Solnica-Krezel L, Stainier DY, Driever W (1994) Implications for dorsoventral axis determination from the zebrafish mutation janus. *Nature* 370:468–471
75. Stainier DY, Fouquet B, Chen JN, Warren KS, Weinstein BM et al (1996) Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. *Development* 123:285–292
76. Talbot WS, Schier AF (1999) Positional cloning of mutated zebrafish genes. *Methods Cell Biol* 60:259–286
77. Brownlie A, Donovan A, Pratt SJ, Paw BH, Oates AC et al (1998) Positional cloning of the zebrafish sauternes gene: a model for congenital sideroblastic anaemia. *Nat Genet* 20:244–250
78. Henikoff S, Till BJ, Comai L (2004) TILLING. Traditional mutagenesis meets functional genomics. *Plant Physiol* 135:630–636
79. Amsterdam A, Burgess S, Golling G, Chen W, Sun Z et al (1999) A large-scale insertional mutagenesis screen in zebrafish. *Genes Dev* 13:2713–2724
80. Chen W, Burgess S, Golling G, Amsterdam A, Hopkins N (2002) High-throughput selection of retrovirus producer cell lines leads to markedly improved efficiency of germ line-transmissible insertions in zebra fish. *J Virol* 76:2192–2198
81. Golling G, Amsterdam A, Sun Z, Antonelli M, Maldonado E et al (2002) Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. *Nat Genet* 31:135–140
82. Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* 26:695–701
83. Meng A, Tang H, Yuan B, Ong BA, Long Q et al (1999) Positive and negative cis-acting elements are required for hematopoietic expression of zebrafish GATA-1. *Blood* 93:500–508
84. Torgersen J, Collas P, Alestrom P (2000) Gene-gun-mediated transfer of reporter genes to somatic zebrafish (*Danio rerio*) tissues. *Mar Biotechnol* (NY) 2:293–300
85. Powers DA, Hereford L, Cole T, Chen TT, Lin CM et al (1992) Electroporation: a method for transferring genes into the gametes of zebrafish (*Brachydanio rerio*), channel catfish (*Ictalurus punctatus*), and common carp (*Cyprinus carpio*). *Mol Mar Biol Biotechnol* 1:301–308
86. Halloran MC, Sato-Maeda M, Warren JT, Su F, Lele Z et al (2000) Laser-induced gene expression in specific cells of transgenic zebrafish. *Development* 127:1953–1960
87. Huang CJ, Jou TS, Ho YL, Lee WH, Jeng YT et al (2005) Conditional expression of a myocardium-specific transgene in zebrafish transgenic lines. *Dev Dyn* 233:1294–1303
88. Linney E, Hardison NL, Lonze BE, Lyons S, DiNapoli L (1999) Transgene expression in zebrafish: a comparison of retroviral-vector and DNA-injection approaches. *Dev Biol* 213:207–216
89. Linney E, Udvadia AJ (2004) Construction and detection of fluorescent, germline transgenic zebrafish. *Methods Mol Biol* 254:271–288
90. Bogers R, Mutsaers E, Druke J, De Roode DF, Murk AJ et al (2006) Estrogenic endpoints in fish early life-stage tests: luciferase and vitellogenin induction in estrogen-responsive transgenic zebrafish. *Environ Toxicol Chem* 25:241–247
91. Ashworth R, Brennan C (2005) Use of transgenic zebrafish reporter lines to study calcium signalling in development. *Brief Funct Genomic Proteomic* 4:186–193
92. Higashijima S, Masino MA, Mandel G, Fetcho JR (2003) Imaging neuronal activity during zebrafish behavior with a genetically encoded calcium indicator. *J Neurophysiol* 90:3986–3997
93. Mattingly CJ, McLachlan JA, Toscano WA Jr (2001) Green fluorescent protein (GFP) as a marker of aryl hydrocarbon receptor (AhR) function in developing zebrafish (*Danio rerio*). *Environ Health Perspect* 109:845–849
94. Higashijima S, Hotta Y, Okamoto H (2000) Visualization of cranial motor neurons in live transgenic zebrafish expressing green fluorescent protein under the control of the islet-1 promoter/enhancer. *J Neurosci* 20:206–218
95. Hill A, Howard CV, Strahle U, Cossins A (2003) Neurodevelopmental defects in zebrafish (*Danio rerio*) at environmentally relevant dioxin (TCDD) concentrations. *Toxicol Sci* 76:392–399
96. Blechinger SR, Warren JT Jr, Kuwada JY, Krone PH (2002) Developmental toxicology of cadmium in living embryos of a stable transgenic zebrafish line. *Environ Health Perspect* 110:1041–1046
97. Amanuma K, Takeda H, Amanuma H, Aoki Y (2000) Transgenic zebrafish for detecting mutations caused by compounds in aquatic environments. *Nat Biotechnol* 18:62–65
98. Scalzo FM, Levin ED (2004) The use of zebrafish (*Danio rerio*) as a model system in neurobehavioral toxicology. *Neurotoxicol Teratol* 26:707–708
99. Rodriguez F, Lopez JC, Vargas JP, Broglio C, Gomez Y et al (2002) Spatial memory and hippocampal pallium through vertebrate evolution: insights from reptiles and teleost fish. *Brain Res Bull* 57:499–503
100. Gerlai R (2003) Zebra fish: an uncharted behavior genetic model. *Behav Genet* 33:461–468
101. Carvan MJ 3rd, Loucks E, Weber DN, Williams FE (2004) Ethanol effects on the developing zebrafish: neurobehavior and skeletal morphogenesis. *Neurotoxicol Teratol* 26:757–768
102. Giacomini NJ, Rose B, Kobayashi K, Guo S (2006) Antipsychotics produce locomotor impairment in larval zebrafish. *Neurotoxicol Teratol* 28:245–250
103. Levin ED, Swain HA, Donerly S, Linney E (2004) Developmental chlorpyrifos effects on hatchling zebrafish swimming behavior. *Neurotoxicol Teratol* 26:719–723
104. Bretau S, Lee S, Guo S (2004) Sensitivity of zebrafish to environmental toxins implicated in Parkinson's disease. *Neurotoxicol Teratol* 26:857–864
105. Samson JC, Goodridge R, Olobatuyi F, Weis JS (2001) Delayed effects of embryonic exposure of zebrafish (*Danio rerio*) to methylmercury (MeHg). *Aquat Toxicol* 51:369–376
106. Kokel D, Bryan J, Laggner C, White R, Cheung CY et al (2010) Rapid behavior-based identification of neuroactive small molecules in the zebrafish. *Nat Chem Biol* 6:231–237

107. Corredor-Adamez M, Welten MC, Spaink HP, Jeffery JE, Schoon RT et al (2005) Genomic annotation and transcriptome analysis of the zebrafish (*Danio rerio*) hox complex with description of a novel member, hox b 13a. *Evol Dev* 7:362–375
108. Lo J, Lee S, Xu M, Liu F, Ruan H et al (2003) 15000 unique zebrafish EST clusters and their future use in microarray for profiling gene expression patterns during embryogenesis. *Genome Res* 13:455–466
109. Linney E, Dobbs-McAuliffe B, Sajadi H, Malek RL (2004) Microarray gene expression profiling during the segmentation phase of zebrafish development. *Comp Biochem Physiol C Toxicol Pharmacol* 138:351–362
110. van der Ven K, De Wit M, Keil D, Moens L, Van Leemput K et al (2005) Development and application of a brain-specific cDNA microarray for effect evaluation of neuro-active pharmaceuticals in zebrafish (*Danio rerio*). *Comp Biochem Physiol B Biochem Mol Biol* 141:408–417
111. Mathavan S, Lee SG, Mak A, Miller LD, Murthy KR et al (2005) Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS Genet* 1:260–276
112. Clark MD, Hennig S, Herwig R, Clifton SW, Marra MA et al (2001) An oligonucleotide fingerprint normalized and expressed sequence tag characterized zebrafish cDNA library. *Genome Res* 11:1594–1602
113. Handley-Goldstone HM, Grow MW, Stegeman JJ (2005) Cardiovascular gene expression profiles of dioxin exposure in zebrafish embryos. *Toxicol Sci* 85:683–693
114. Ton C, Stamatou D, Liew CC (2003) Gene expression profile of zebrafish exposed to hypoxia during development. *Physiol Genomics* 13:97–106
115. Hoyt PR, Doktycz MJ, Beattie KL, Greeley MS Jr (2003) DNA microarrays detect 4-nonylphenol-induced alterations in gene expression during zebrafish early development. *Ecotoxicology* 12:469–474
116. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63
117. Mandrell D, Moore A, Jephson C, Sarker M, Lang C et al (2011) Automated zebrafish chorion removal and single embryo transfer: optimizing throughput of zebrafish developmental toxicity screens. *J Lab Autom* (submitted)
118. Wang W, Liu X, Gelinas D, Ciruna B, Sun Y (2007) A fully automated robotic system for microinjection of zebrafish embryos. *PLoS One* 2:e862
119. Carpenter AE (2007) Image-based chemical screening. *Nat Chem Biol* 3:461–465
120. Mayr LM, Fuerst P (2008) The future of high-throughput screening. *J Biomol Screen* 13:443–448
121. Milan DJ, Peterson TA, Ruskin JN, Peterson RT, MacRae CA (2003) Drugs that induce repolarization abnormalities cause bradycardia in zebrafish. *Circulation* 107:1355–1358
122. Berghmans S, Butler P, Goldsmith P, Waldron G, Gardner I et al (2008) Zebrafish based assays for the assessment of cardiac, visual and gut function—potential safety screens for early drug discovery. *J Pharmacol Toxicol Methods* 58:59–68
123. Grassian VH (2008) When size really matters: size-dependent properties and surface chemistry of metal and metal oxide nanoparticles in gas and liquid phase environments†. *J Phys Chem C* 112:18303–18313
124. MacPhail RC, Brooks J, Hunter DL, Padnos B, Irons TD et al (2009) Locomotion in larval zebrafish: influence of time of day, lighting and ethanol. *Neurotoxicology* 30:52–58
125. Usenko CY, Harper SL, Tanguay RL (2007) *In vivo* evaluation of carbon fullerene toxicity using embryonic zebrafish. *Carbon* 45:1891–1898
126. Harper SL, Dahl JL, Maddux BLS, Tanguay RL, Hutchison JE (2008) Proactively designing nanomaterials to enhance performance and minimize hazard. *I J Nanotechnol* 5:124–142
127. Holdsworth DW, Thornton MM (2002) Micro-CT in small animal and specimen imaging. *Trends Biotechnol* 20:S34–S39
128. Harper SL, Usenko C, Hutchinson JE, Maddux BLS, Tanguay RL (2008) *In vivo* biodistribution and toxicity depends on nanomaterial composition, size, surface functionalization and route of exposure. *J Exp Nanosci* 3:195–206
129. Magrez A, Kasas S, Salicio V, Pasquier N, Seo JW et al (2006) Cellular toxicity of carbon-based nanomaterials. *Nano Lett* 6:1121–1125
130. Nel A, Xia T, Madler L, Li N (2006) Toxic potential of materials at the nanolevel. *Science* 311:622–627
131. Colvin V (2003) The potential environmental impact of engineered nanomaterials. *Nat Biotechnol* 21:1166–1170
132. Jiang W, KimBetty YS, Rutka JT, ChanWarren CW (2008) Nanoparticle-mediated cellular response is size-dependent. *Nat Nano* 3:145–150

Natural Gas Power

RAUB W. SMITH, S. CAN GÜLEN

GE Infrastructure-Energy, Schenectady, NY, USA

Article Outline

Glossary
 Defining the Subject
 Introduction
 Gas Turbine Power Plants
 Future Directions
 Bibliography

Glossary

Brayton cycle The thermodynamic cycle describing the operation of a gas turbine. In a combined cycle, it is the *topping* cycle due to its relative position vis-à-vis Rankine cycle on a temperature-entropy surface.

Carnot cycle Also known as the Carnot engine, it is the embodiment of the second law of thermodynamics in the form of a theoretical cycle comprising two isentropic and two isothermal processes. No heat engine operating in a thermodynamic cycle can be more efficient than the corresponding Carnot engine defined by the constant mean-effective heat addition and heat rejection temperatures.

Cogeneration See combined heat and power (CHP).

Combined cycle power plant A fossil-fired power plant that combines two types of prime movers, usually one or more gas turbines and one or more steam turbines (STs), whose operation are governed by their respective thermodynamic cycles, i.e., Brayton and Rankine.

Combined heat and power (CHP) The term used for fossil-fired power plants, which, in addition to their primary product, electric power delivered to the grid, also supply a secondary product in terms of useful thermal energy.

Combustor A mechanical device to facilitate controlled mixing and reaction of an oxidizer (in almost all cases air) and a fuel (in almost all cases a pure hydrocarbon or a mixture thereof in gaseous or liquid phase) to generate high-temperature gaseous product for expansion in a turbine and useful shaft work generation.

Compressor A mechanical device that increases the pressure of a gas by reducing its volume. There are different types of compressors, e.g., axial, radial, and reciprocating, which are suitable to different types of operating regimes.

Efficiency Unless specified otherwise, the thermal efficiency of a power-generating system, which is the dimensionless ratio of generated kWh of electricity to the amount of energy required to generate it. It is the inverse of the heat rate with a suitable conversion factor.

Emissions Gases and solid particles (usually undesirable) released into the air as by-products of a combustion process (e.g., in the boiler of a fossil-fired power plant, gas turbine combustor, or other internal combustion engine) to create electric power or propel a vehicle.

Firing temperature The temperature of the gas turbine combustor exhaust gas at the inlet to the first

stage rotor, which is the starting point of useful shaft work generation.

Gas turbine A prime mover or internal combustion engine comprising a compressor, combustor, and an expander connected via a common shaft, through which air is compressed, burned, and expanded to generate useful shaft work for electric power generation (or thrust in an aircraft jet engine).

Generator A device that converts the mechanical shaft power generated by a prime mover into electrical power.

Global warming The apparent increase in the average temperature of the earth's near-surface air and oceans since the mid-twentieth century and its projected continuation (per *Wikipedia*).

Greenhouse effect The containment of heat from solar radiation striking the earth's surface due to the earth's atmospheric "greenhouse" gases such as carbon dioxide and methane. These gases absorb and emit radiation within the thermal infrared range and are believed to be a primary cause of global warming.

Heat rate Amount of energy required to generate 1 kWh of electricity. It is the inverse of the thermal efficiency with a suitable conversion factor.

Heating value The thermal energy produced by completely burning a unit mass of fuel in a combustor to produce carbon dioxide and water. If the water is in a gaseous phase, the heating value is referred to as net or lower heating value (LHV). If the water is in a liquid phase, the heating value is referred to as gross or higher heating value (HHV).

Heat recovery steam generator (HRSG) Also known as the heat recovery boiler (HRB), HRSG is a cross-flow tubular heat exchanger that recovers the exhaust heat from a prime mover (e.g., a gas turbine) and produces steam at high pressure and temperature that is used in a steam turbine (ST) for additional power generation. HRSG is the key equipment that "combines" gas and steam turbines in a combined cycle power plant.

Rankine cycle The thermodynamic cycle describing the operation of a steam turbine. In a combined cycle, it is the *bottoming* cycle due to its relative position vis-à-vis Brayton cycle on a temperature-entropy surface.

Steam turbine A prime mover or the power-generating part of an external combustion engine comprising one or more sections connected via a common shaft, through which steam flows, expands, and discharges to a condenser to generate useful shaft work for electric power generation or propulsion.

Defining the Subject

Natural gas is an important fossil fuel that has played an increasingly significant role in worldwide electric power generation since the 1980s. The key driver underlying the importance of natural gas as a vital enabler of modern living has been its relative advantage vis-à-vis other fossil fuels in terms of emissions and pollutants. In comparison to coal, the primary fossil fuel used for electric power generation in the world on a constant consumption basis, natural gas emits nearly 45% less CO₂, 80% less nitrogen oxides (NO_x) with negligible amounts of sulfur oxides, particulates, and mercury.

The environmental advantages of natural gas are further amplified by the significantly higher thermal efficiency of the power plants that burn it for electric power generation in comparison to other variants of oil or coal. Natural gas burning modern gas turbines can readily reach efficiencies of 56–57% in combined cycle configurations. This is well above the average efficiency of existing coal burning steam turbine plants (e.g., low 30s) as well as the most advanced ultra-supercritical designs thereof (e.g., low 40s).

With the beginning of the twenty-first century, economic projections pointed to natural gas as the fastest growing fuel for worldwide electric power generation, with an expected average annual rate of growth of about 3.7% between 2005 and 2030. During this period, natural gas will account for almost one-fourth of the world's total net electricity generation [1]. While such long-term forecasts are highly susceptible to uncertainties such as the impact of potential governmental policies or legislation limiting the use of fossil fuels in favor of nuclear and renewable sources of energy (largely driven by concerns over global warming), it is almost a foregone conclusion that natural gas will continue to be an important part of the world's energy and electric power generation portfolio in the foreseeable future.

Given this impact of natural gas on future electrical production, it is important to establish a basic understanding of the history, current state of the art and future possibilities of power generation using natural gas combustion in order to have a complete picture of future energy sustainability.

Introduction

Natural gas is a fossil fuel comprising hydrocarbons, primarily methane (CH₄) up to 90% on a volumetric basis. It differs from the other fossil fuels in that it is naturally available as a gas (hence the name) as opposed to a solid (i.e., coal) or liquid (i.e., oil). Like other fossil fuels, the basic mechanism (thermogenic) of natural gas formation involves the extremely prolonged high-pressure compression of organic matter remains (e.g., plants, animals that lived millions of years ago on earth). The gaseous phase is the result of extremely high temperatures associated with the depth of the organic matter, typically 1–2 miles below the earth's crust. At even deeper levels, natural gas consists almost entirely of pure methane.

Another (biogenic) mechanism for forming natural gas is the chemical transformation of organic matter by microorganisms. The resulting gas contains methane and carbon dioxide. A well-known example of this mechanism is the landfill gas that results from the decomposition of waste materials deposited into large landfills. Other biogas formation mechanisms are sludge digestion in the tanks of sewage treatment plants (sewage gas) and anaerobic fermentation of agricultural waste. Biogas formation takes place near the earth's surface in the absence of oxygen and the resulting gas usually leaks to atmosphere. Controlled capture of landfill gas and other biogases for utilization in electric power generation is a promising technology. (Uncontrolled energy release (explosion) from the trapped landfill gas leading to death and destruction in the poverty-stricken areas of the world is an unfortunate event that appears in the news from time to time.) Biogas utilization is also important because the main constituent, methane, itself is a principal greenhouse gas (GHG). In fact, methane is nearly 20 times more effective than CO₂ in trapping heat emanating from the earth's surface. Therefore, combustion of methane (with resultant CO₂ emissions) that otherwise

would leak into the atmosphere is a preferable trade-off from GHG reduction point of view.

Coal mine gas is another type of natural gas that is released during pit coal mining. Depending on the time of capture and method, there are different types. The most common types are coal bed methane (CBM) and coal mine methane (CMM). The former is largely pure methane and its production is independent of coal mining. The latter is released during active mining and presents a significant danger of violent explosion to miners. Proper ventilation of mine shafts to remove and vent CMM into atmosphere is essential for the safety of mining operation. The methane content of CMM is variable (25–60%) and can change suddenly, which might be problematic for its use as a fuel. A third type of coal mine gas is abandoned mine methane (AMM), which can seep from mines that are no longer active. Annual coal mine methane emissions account for about 8% of total anthropogenic methane emissions or about 28 billion m³ of carbon dioxide equivalent [3]. Utilization of CMM and AMM for power generation is critical to the prevention of release of a GHG into atmosphere.

Energy content of the natural gas is measured in terms of Btu (kJ or kcal) per lb (kg) or per standard cubic feet (cubic meters). In many practical calculations, natural gas is assumed to be 100% pure methane, which has an energy content of 21,515 Btu/lb (50,044 kJ/kg or 11,953 kcal/kg) on a lower heating value (LHV) basis or 914 Btu/ft³ (34,050 kJ/m³ or 8,133 kcal/m³) at normal (standard) conditions (herein defined as 77°F (25°C) and 1 atm). The actual composition may vary depending on the source and

treatment. Based on the mean of over 6,800 samples of pipeline quality natural gas taken in 26 major metropolitan areas of the USA [4], the methane content is 93.1% with a heating value of 930 Btu/ft³ (8,275 kcal/m³). A sampling of composition and heating value of most common natural gas variants is given in Table 1.

Production and distribution companies measure natural gas in volumetric terms, i.e., in multiples of cubic feet such as millions of cubic feet, trillions of cubic feet, etc. In power generation calculations, where natural gas is the fuel burned in the combustor of the gas turbine, the more practical approach is to use the mass flow rate (lb/s or kg/s) along with the LHV value or the product of the two, which is referred to as heat consumption (MMBtu/h or kW).

The purpose of a power plant is to generate electric power. Therefore, the key performance metric of an electric power generator is the net thermal efficiency, which is the ratio of the electric power measured at the generator terminals (after subtracting all the power that is spent to keep the plant running) to the plant's fuel energy consumption. The most efficient way to burn natural gas for electric power generation is in a gas turbine combined cycle (GTCC) power plant. Today's state-of-the-art CC power plant has a rated net efficiency of 58% on an LHV basis. In other words, for each 1 lb/s of natural gas (about 23.5 ft³/s as 100% methane) that enters the combustor of the gas turbine, (using the formula $58\% \times 21,515 \text{ Btu/lb} \times 1 \text{ lb/s} \times 1.05506 \text{ kW/Btu} = 13,166 \text{ kW}$) about 13.2 MW of electric power (net) is delivered to the grid. This is enough power to sustain nearly 13,000

Natural Gas Power. Table 1 Composition and heating values of typical natural gases

Component (%)	Coal bed methane	Coal mine methane	Landfill gas	Biogas	Average NG
O ₂	2.1	12.6	0–1	0–2	
N ₂	8.2	46.8	2–5	0–10	1.6
CH ₄	85.9	40.0	45–60	50–75	93.1
CO ₂			40–65	25–50	1.0
C _n H _{2n}	3.8				4.3
LHV, kcal/nm ³ (Btu/scf)	7,762	3,200	3,350–4,775	3,580–5,975	8,275
	872	360	375–535	400–670	930

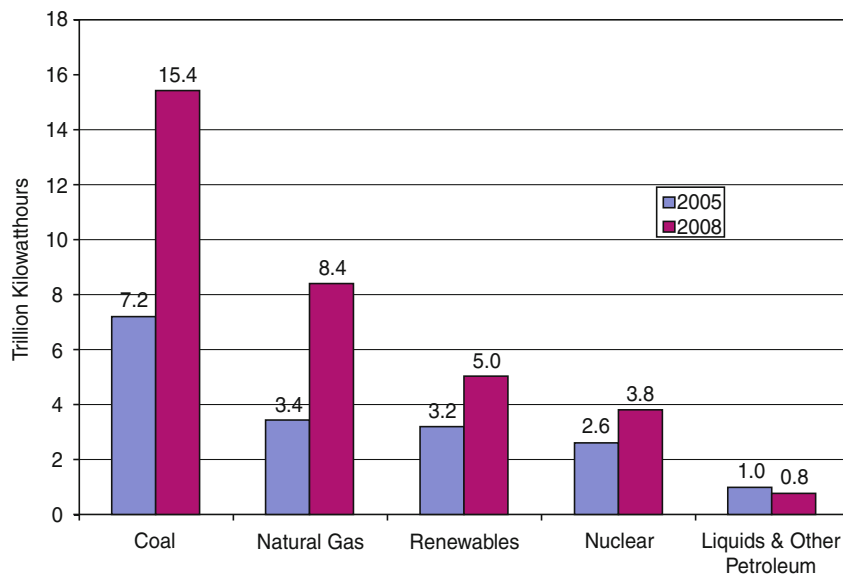
average homes for 1 h at an average power consumption of 1 kW.

In 2005, natural gas accounted for 3.4 trillion kWh of net electricity production for about 20% of the world total (see Fig. 1). This contribution is projected to reach 8.4 trillion kWh in 2030 or about 25% of the world total corresponding to a total natural gas consumption of 55.5 trillion ft³ (i.e., 35% of total world production of 158.6 trillion ft³). Using the calculation above as a rough guide, this translates into an average natural gas-fired power plant thermal efficiency of 56.4%. The natural gas reserves of the world are estimated at about 6,186 trillion ft³ with top 20 countries contributing nearly 90% of it (see Fig. 2). When one considers the uncertainty in the cited projections, it appears safe to say that there are ample deposits of natural gas to sustain electric power production well into the twenty-first century.

As mentioned in the beginning, natural gas is a superior fossil fuel in terms of emissions and pollutants vis-à-vis other fossil fuels, i.e., coal and oil. This is not to say that natural gas is totally harmless, especially when one considers the emissions of CO₂ – the most abundant anthropogenic greenhouse gas in the atmosphere and, as such, the main culprit suspected in driving global warming. A modern gas turbine

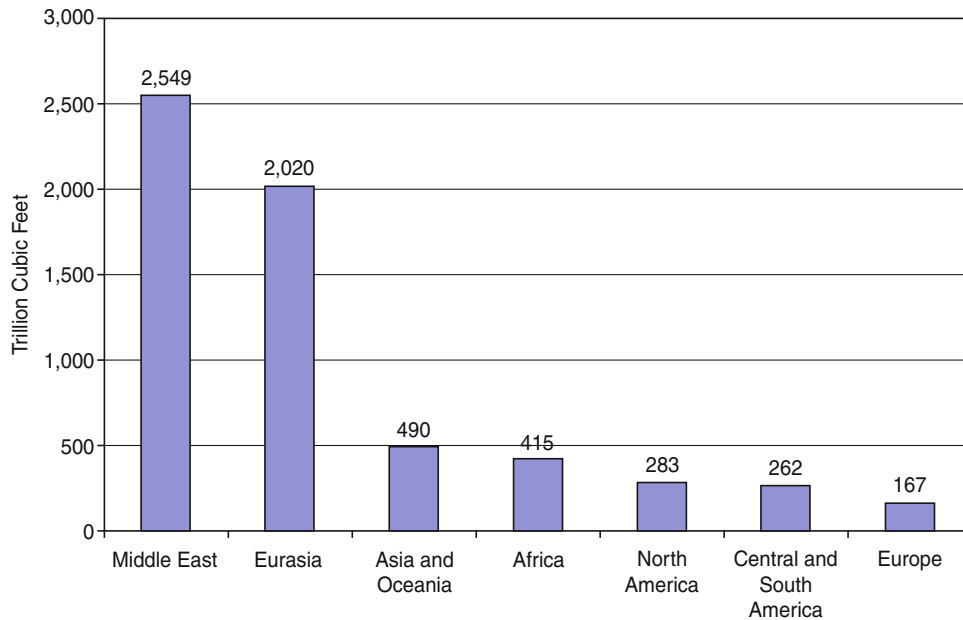
combustor generates 2.75 lbs of CO₂ per lb of natural gas burned. In other useful relative terms, this is approximately 129,000 lb (58,500 kg) of CO₂ per billion Btu (1.06 million MJ) of energy input or 1,125 lb/MWh generated by a gas turbine (GT) in simple cycle configuration (cf., 750 lb in combined cycle configuration). Since not all power generation equipment performs at today's higher efficiency level, real averages would be higher, e.g., 1,300 and 800 for simple and combined cycles, respectively. For the projected natural gas-based electric power generation in 2030 cited above, at a rate of 1,300 lb/MWh, this translates into 5 billion metric tons of CO₂ released into earth's atmosphere. (In order to get an idea about the magnitude of this number, consider that the estimated weight of carbon dioxide in the atmosphere was about 2,163 billion metric tons before the industrial revolution.) This is approximately half of the total 2030 projection of CO₂ emission from natural gas combustion, which itself is about 20% of the 42.3 billion metric tons for all fossil fuels.

While an important contributor to the greenhouse effect, natural gas is extremely advantageous from the perspective of two other significant combustion-related pollution phenomena, i.e., smog and acid rain. Smog (whose primary constituent is ozone) is formed



Natural Gas Power. Figure 1

World Net Electricity Generation. Renewables include hydroelectric (Source: Energy Information Administration (EIA) [1])



Natural Gas Power. Figure 2

World Natural Gas Reserves, January 2008. Asia, Oceania and Europe exclude countries that were part of the former USSR whereas Eurasia includes only those. Source: Energy Information Administration (EIA) [2]

by a chemical reaction of carbon monoxide, nitrogen oxides, volatile organic compounds, and heat from sunlight [5]. Acid rain is formed when sulfur dioxide and nitrogen oxides react with water vapor and other chemicals in the presence of sunlight to form various acidic compounds in the air [5]. Natural gas combustion products are virtually free of the two main culprits contributing to these severe pollution problems, namely, sulfur oxides (SO_x) and particulate matter, and they have 80% less NO_x than products of coal combustion.

In general (notwithstanding some future fuel cell designs discussed later), the most efficient means to utilize natural gas as a fossil fuel in electric power generation is to burn it in internal combustion (IC) engines. The two most widely used IC engines in electric power generation are gas turbine and reciprocating internal combustion or diesel engine. (Note that the term *diesel* in this context refers to the thermodynamic cycle of the compression-ignition (CI) IC engine and not the fuel.) While possible in theory, natural gas is not used with external combustion technology (e.g., boiler and steam turbine plants) due to the pronounced superiority of the IC engines in efficiency,

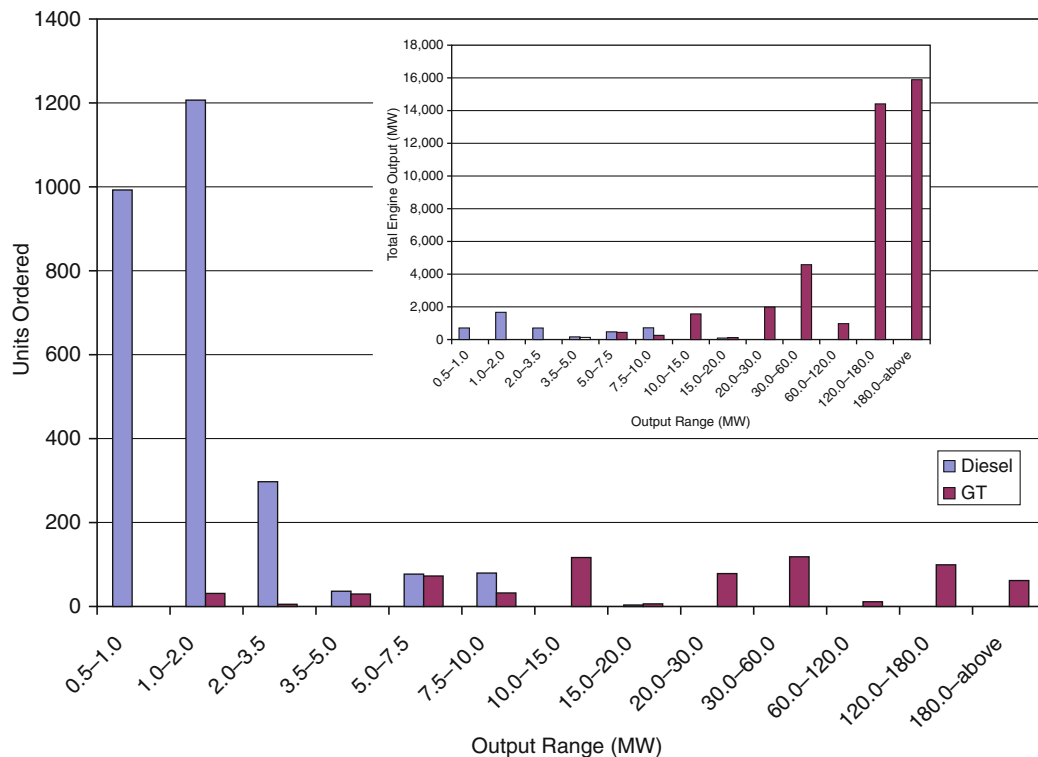
emissions, power density (i.e., favorable cost vs. size trade-off) and operational flexibility.

Modern gas turbines are large air-breathing turbomachines with extremely large power output. For example, consider a 50-Hz (3,000 rpm) 300+ MW net power output unit as listed in a 2008 trade publication [6]. This machine ingests air at ISO conditions (15°C and 1 atm at 60% relative humidity) at a rate of nearly 1,550 lb/s (705 kg/s), compresses it to a pressure that is 18 times that of the ambient and combusts it with about 35 lb/s (16 kg/s) of natural gas (100% CH_4) generating 312 MW net electric power for a net thermal efficiency of 39.3%. At the inlet to the expander section of the gas turbine, where they produce useful shaft work, the combustion products are at nearly 1,500°C (2,732°F). This is well above the melting point of the most advanced superalloy materials that are used in the manufacture of turbine expander components. In order to ensure the survival of the turbine parts under those extreme conditions for thousands of hours of continuous operation, nickel-based superalloy components are protected by thermal barrier coatings (TBC) and internally cooled by using the “cold” air extracted from the turbine compressor. (Consider that

without cooling and TBC, the first stage vanes of a modern GT would survive barely ten seconds before melting away.) Utilizing those and other technologies, some of which are adopted from the advanced military and civilian aircraft engine–related research and development, land-based heavy-duty industrial gas turbines are true marvels of human engineering.

Diesel or compression-ignition IC engines are no slouches themselves. In fact, the efficiency of a tri-fuel (one that can burn natural gas, light and heavy fuel oils) unit rated at nearly 17 MW net electric output is 47.3% on an LHV basis when operating with natural gas fuel [7]. They are eminently suitable to landfill, biogas, and coal mine gas applications. They are at a disadvantage because of their relatively low power density and cost. The sample multifuel diesel engine cited above has a power density of 0.05 kW/kg of engine weight or 215 kW/m² of engine footprint. These numbers compare very unfavorably to the power density of a heavy-duty industrial gas turbine. For a typical advanced

machine with the performance similar to the sample cited above, the power density is about 0.8 kW/kg of engine weight or about 3,000 kW or more per m² of engine footprint. In order to appreciate the power density of a GT vis-à-vis renewable energy–based systems, consider that the power density of typical wind and solar radiation are 0.4–0.8 (depending on altitude and speed) and 1.4 kW/m², respectively [8]. One would need hundreds of acres of wind turbine farms or solar collector fields to replace a single advanced F-Class gas turbine. This gross disparity in the power density of the two IC engine technologies is reflected in the annual unit order numbers displayed in Fig. 3. In terms of the numbers of units ordered in 2005–2006 [9], reciprocating IC engines dominate (nearly 2,700 units vs. 666 for gas turbines) but the significant bulk of the orders are for units rated at 3.5 MW or less. In terms of the total MW rating of the orders, all dual-fuel and natural gas diesel engines add up to 4.4 GW as opposed to more than 40 GW for all gas turbines.



Natural Gas Power. Figure 3

Dual-fuel and natural gas diesel engine and gas turbine orders, June 2005 through May 2006 [9]

At this point, it should be fairly obvious that a treatise on the subject of electric power generation using natural gas is essentially a treatise on modern land-based or heavy-duty industrial gas turbines. This is especially true for the past and the present of the technology. The future of electric power generation using natural gas as a fuel source might have a place for the fuel cell technology (with or without gas turbines). Ongoing research activities include solid oxide (SOFC), molten carbonate (MCFC), and polymer electrolyte (PEFC) technologies. From where one stands today, the most promising use for these technologies seems to be found in the fields of transportation and portable power. With the exception of distributed power generation, in which small-scale (1–500 kW) units power individual residences, office buildings, etc., the prospect of fuel cell-based technologies to replace internal combustion engines and specifically gas turbines as large-scale base load power-generating units in the near term, especially in a cost-effective manner, is practically nil.

A Brief History of Natural Gas

Natural gas has been around for quite a while. In China, nearly 2,500 years ago, early innovators used natural gas escaping naturally from the ground to light their temples and heat brine for distillation. Bamboo sections, split lengthwise, were glued together, and bound with twine, allowing the gas to be transported. In the early seventeenth century, French explorers near Lake Erie reported seeing natives burning gas naturally seeping from the ground. The first industrial use of natural gas can be traced back to England in the late eighteenth century when gas produced from coal was used to light houses and streets. In the USA, the first natural gas well was dug in 1821 in Fredonia, NY, four decades before Colonel Drake hit upon oil and gas at a depth of 69 feet in Lake Erie, PA. The well, dug by William Hart, was 27 ft deep in a creek bed. The extracted gas was transported through hollow log pipes to Fredonia for use in street lamps. Throughout most of the nineteenth century, natural gas was used almost exclusively for the same purpose, i.e., lighting. Following the invention of the Bunsen burner in 1885, natural gas also presented households with a more convenient means of cooking and heating. The main impediment

to a wider utilization of natural gas was the lack of a transportation mechanism. The primary method, still in use today, is through an underground pipeline. While the first significantly long pipeline was constructed in 1891 (120 miles long between central Indiana and Chicago), it was only after World War II when technology advances made construction of longer and more reliable pipelines feasible. Until that time, natural gas was either allowed to vent into the atmosphere (intentionally or unintentionally) or burned in flares (as a by-product of coal mining and oil exploration) or just left in the ground when found alone. In fact, when the first astronauts looked at earth from the orbit, the most prominent sights on the earth's surface were the flares from the oil fields in Middle East and Africa. Even now about 150 million m³ of natural gas, roughly equivalent to annual US residential consumption, is flared to the atmosphere. The contribution of the flared natural gas (NG) to global GHG emissions is about 400 million metric tons, equivalent to nearly one third of the annual US car exhaust emissions.

Today natural gas is used for much more than street lighting or household cooking. In fact, residential use of natural gas accounts for only a fifth of the total US consumption, with more than 60% used by the industrial sector including electric power generation. Until the end of 1980s, high-efficiency fossil fuel (mainly coal) firing or nuclear-powered steam turbine-based power plants supplied the base load electric generation capacity (6,000–8,000 h/year). When newly designed high-efficiency steam plants entered the service, older less-efficient units were relegated to intermediate duty operation (e.g., 4,000 h/year). Early generation, low-efficiency gas turbines, along with old steam plants and hydro pumped-storage plants, were largely limited to short-duration, peak-shaving-type operations. Due to changing economic parameters, introduction of protective environmental regulations, and major technological advances in gas turbine technology, starting in mid-1980s and early 1990s, natural gas became the fuel of choice for new power plants. The steady increase in coal-fired steam plant efficiencies leveled off at the low 40% range, whereas new “F-Class” gas turbines with high firing temperatures, almost exclusively in combined cycle configuration, opened the door for thermal efficiencies easily exceeding 50% and even pushing for 60%. Recently, especially during the so-called energy

boom of early 2000s, a large fraction of new capacity addition in the USA has been in natural gas-fired units, e.g., nearly 95% or 22 GW in 2000. Since then, this rate has tapered off significantly. Natural gas-fired power plant capacity addition in terms of construction starts totaled about the same, i.e., 22 GW, in the years 2005–2007 [10]. Scheduled construction start is about 9 GW for 2010, about the same as in 2009 but lower than the peak year of 2008 with nearly 12 GW. As pointed out earlier, natural gas is projected to be the fastest growing fuel for worldwide electric power generation. Thus, the pace of construction and commissioning of plants burning natural gas might be expected to pick up again. Construction of natural gas-fired power plants scheduled to start between 2011 and 2015 was estimated to be 44.5 GW [10]. Even though some projects will undoubtedly be cancelled or postponed, this is a strong positive sign for the future role of natural gas in electric power production.

The driver of the ascendancy of natural gas as the star of fossil fuels is the gas turbine, which also plays a key role in its transportation and distribution via huge pipeline networks. Heavy-duty industrial and/or aeroderivative gas turbines drive the large centrifugal gas compressors, which increase the pressure of natural gas up to 1,500 psia (~ 100 bar). In most cases, these gas turbines burn the natural gas taken from the pipeline. Additional compressor stations, placed at regular intervals along the hundreds of miles long pipelines, maintain line pressures reduced via friction between pipe and gas, which can reach flow velocities up to 40 km/h.

A Brief History of Gas Turbine

When lieutenant colonel Heinz Bär died in a civilian aircraft accident in 1956 in (then) West Germany, he was the world's leading jet fighter ace with 16 victories to his credit. Some five decades after his death and more than six decades after his last aerial exploits, he is still unrivaled. The machine that enabled Bär to permanently enter the annals of military history was the world's first mass production jet fighter, the Messerschmitt Me-262. It was powered, remarked another famous German fighter pilot, "as if pushed by angels" – actually two 1984-lb thrust Jumo-004 gas turbine jet engines.

Developed under the leadership of Franz Anselm, the Jumo-004 was essentially the engine that opened up the jet age. Quite naturally, Dr. Anselm did not conceive Jumo-004 in a vacuum. Paraphrasing Sir Isaac Newton, he was standing on the shoulders of giants such as Sir Frank Whittle and Hans von Ohain, who are widely recognized as the pioneers of modern turbojet engine. Sir Frank and Dr. Ohain themselves almost certainly acknowledged early inventors such as Barber (first GT patent in 1791), Stolze (first GT to be built and tried in 1900–1904), Holzwarth, and others, who showed them the path to the ultimate gas turbine engine design through their patents on air and steam turbines.

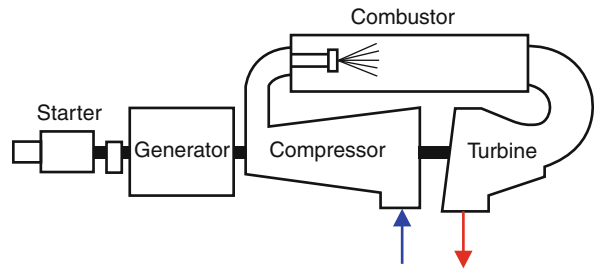
It is futile to attempt to plot a straight historical line from Hero's turbine (which had no purpose other than entropy generation) to today's modern 2,600°F firing temperature class steam-cooled H-System™. Countless inventors laid down the building blocks that eventually led to the modern gas turbine engine. For an excellent short history of turbomachinery in general and gas turbines in particular, the reader should consult the relevant chapter in Wilson and Korakianitis [11], who highlight the contributions of lesser known (unfairly) inventors such as Aegidius Elling and George Jendrassik. Another good narrative with an emphasis on turbojets and German research and development can be found in Hans von Ohain's foreword to Mattingly's book on gas turbine propulsion [12]. For a detailed look at turbojet development activities of Sir Frank Whittle and Hans von Ohain in their respective countries, three articles by Meher-Homji can be consulted [13–15].

Quite obviously, wars or the prewar atmosphere of the western world in the first half of the last century pushed the development of gas turbines primarily as a military aircraft propulsion device. Starting in the late 1940s and early 1950s, engineers in USA and UK managed to derive robust, land-based power generation and, later, marine propulsion engines from these aircraft designs. Claire Soares essentially came to the conclusion that the involvement of so many people from so many different countries with immense pride in their work led to selective "histories" favoring selected milestones [16]. Her brief summary can be consulted for alternate takes on the development history of gas turbines. For a history of Siemens gas turbine development, a good resource is the paper by Leiste [17].

For a US perspective with an emphasis on GE, the reader should consult Miller and Nemec [18] and Brandt [19]. The paper describing the design and development of the GE MS7001F gas turbine by Brandt [20] is a good reference laying out all design aspects of a modern gas turbine and the trade-offs that are unavoidable in bringing a feasible product into the market.

Even though the development of the modern gas turbine was primarily driven by military and, to a lesser extent, civilian aircraft propulsion considerations, the introduction of stationary (land-based) units dedicated to power production did not lag too far behind. In fact, Aurel Stodola credited Hans Holzwarth with “having built the first economically practical gas turbine [21].” Holzwarth built several types of his “explosion” turbine in the first quarter of the twentieth century. Holzwarth’s turbine was, strictly speaking, a hybrid construction combining the spark-ignition (constant volume) combustion process of an Otto cycle with the axial expansion process of a Brayton cycle. Stodola calculated a thermal efficiency of 25.6% for the test of a 1,500-rpm experimental turbine in Mühlheim–Ruhr in 1919 [21]. Burning a gas with 434 Btu/ft³ heating value the turbine produced about 725 kW. Air and fuel gas was compressed by steam turbine–driven compressors and sequentially injected into the explosion chamber (gas first) at about 30 psia. Stodola reports an average maximum explosion pressure of 160 psia. The combustion products expanded through a two-stage velocity-compounded turbine. Exhaust gas at about 800°F was recovered in an exhaust heat boiler. In that sense, this unit could be considered the first combined cycle. Holzwarth designed and developed several different variants of his turbine between 1907 and 1928. From 1928 on, Brown Boveri Company (BBC) took over the development of Holzwarth turbine and in 1933 installed a blast-furnace gas (BFG)–fired unit in a German steel mill, which was destroyed during World War II.

It is generally accepted that the first commercial stationary gas turbine for electric power generation was erected in Neuchâtel, Switzerland, by the former BBC in 1939, 3 years before the first flight of Me-262 powered by Jumo-004. This fuel oil burning 4 MW machine, primarily used for stand-by and peaking duties, was operational for nearly 70 years (see Fig. 4). The combustion chamber was derived from the



Natural Gas Power. Figure 4

Schematic diagram of Brown Boveri Co.’s Neuchâtel gas turbine (1939)

turbocharged *Velox* boiler, which itself resulted from the BBC work done on the Holzwarth turbine. The BBC turbine was tested under the supervision of Aurel Stodola, who reported an overall thermal efficiency of 17.4%, which is less than half of the efficiency of today’s advanced F-Class machines [22]. This machine was designated by the ASME as a *Historic Mechanical Engineering Landmark* in 1988.

The first gas turbine installed in an electric utility in the USA (Oklahoma Gas & Electric, Belle Isle Station, Oklahoma) was a 3.5 MW GE Frame 3 unit that entered service in 1949. In addition to generating power, the exhaust gas of this gas turbine was utilized to heat the feed water of a conventional steam plant. In other words, the first US electric utility gas turbine was in a “combined” cycle configuration [23]. In fact the combined cycle concept goes back to Emmet’s mercury-vapor process (1925) [24]. For a brief history of the CC power plants and pertinent references, the reader should consult the 1994 Calvin Winsor Rice Lecture by Sir John Horlock [24].

One can look at the technology history of the land-based gas turbine and combined cycle power plants for electric power production in four generations:

1. Generation 1 (1949–1968)
 - (a) Smaller than 30 MW GTs (Frames 3 and 5)
 - (b) Firing temperatures 1,500–1,800°F
 - (c) For repowering and cogeneration
2. Generation 2 (1968–1990)
 - (a) B and E Technology GTs (50–120 MW)
 - (b) Firing temperatures ~2,000°F
 - (c) NO_x emission control using GT water/steam injection or SCR
 - (d) Non-reheat steam cycles

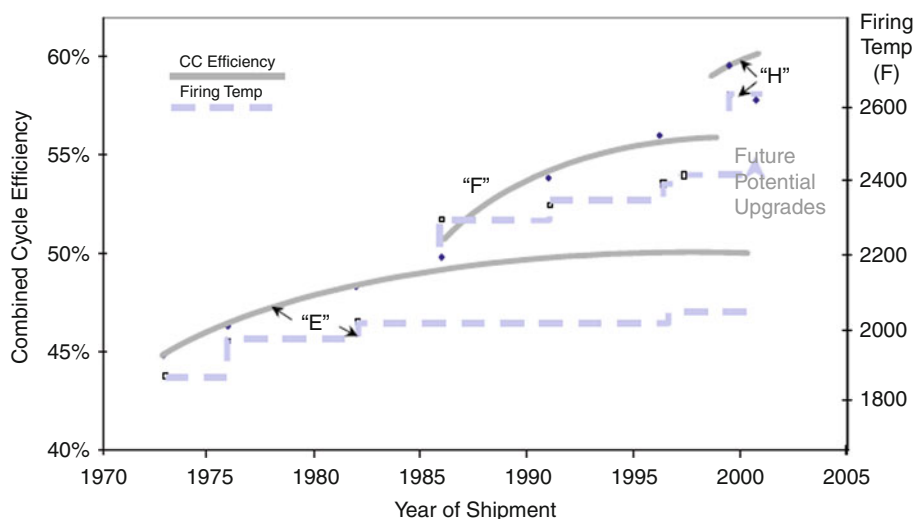
3. Generation 3 (1990–1998)
 - (a) F Technology GTs (75–260 MW)
 - (b) Three air-cooled turbine stages
 - (c) Firing temperatures $\sim 2,400^{\circ}\text{F}$
 - (d) Performance fuel heating (365°F)
 - (e) DLN combustion system for NO_x control
 - (f) Three-pressure, reheat (3PRH) steam cycles
4. Generation 4 (1998–present)
 - (a) H technology GTs (400+ MW in CC)
 - (b) Closed-loop steam cooling (CL-SC) of the first two turbine stages (both stator and rotor); four turbine stages
 - (c) Cooling of cooling air (CAC) for turbine wheel spaces and subsequent stages via a heat exchanger (a kettle reboiler) that generates IP steam to be used in the bottoming cycle
 - (d) Active clearance control (compressor and turbine)
 - (e) Firing temperatures $\sim 2,600^{\circ}\text{F}$
 - (f) Performance fuel heating ($400+^{\circ}\text{F}$)
 - (g) DLN combustion system for NO_x control
 - (h) Three-pressure, reheat (3PRH) steam cycle integrated with the GT Brayton cycle

The historical advances in GT firing temperature and CC efficiency is summarized in Fig. 5. The data illustrates the significant advance in the thermal efficiency of the GT combined cycle power plants, to the

tune of nearly 15 percentage points, which is mainly driven by advances in materials, coatings, and cooling techniques that pushed the firing temperatures by almost 800°F ($\sim 450^{\circ}\text{C}$). In the following chapters, the reader is introduced to the basic thermodynamic principles that govern the performance of natural gas-burning gas turbine power plants. In addition, also provided is a brief overview of state-of-the-art gas turbine power plants, key economic criteria, and operability considerations. Finally, the future directions in gas turbine-based electric power generation systems are briefly elaborated upon.

Gas Turbine Power Plants

Gas turbine in itself or as a key component of a larger power generation system is a subject that can only be covered in dedicated books. Such books do exist and several selected treatises on the subject, which in the opinion of the authors reflect the best available published work in the field, are listed among the bibliography references. The coverage of the subject in the limited space allotted to the current chapter aims to provide the reader with the fundamental thermodynamic considerations that lead from the Carnot cycle to the present-day advanced gas turbine in a logically coherent manner. There is a reason that the gas turbine combined cycle plant is the most efficient heat engine



Natural Gas Power. Figure 5

Incremental evolutions of E-, F- and H-Class GT technology (From Ref. [25])

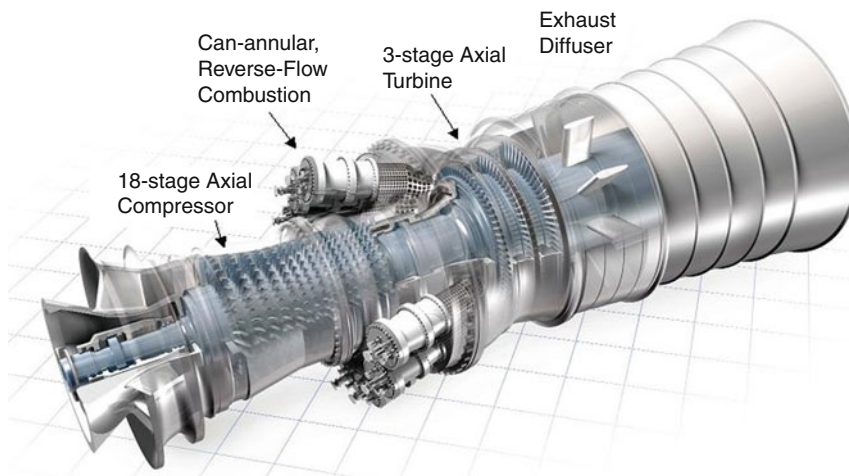
that is commercially available today and, most likely, in the foreseeable future. That reason is embedded in the laws of thermodynamics. Understanding this will enable someone with a basic knowledge of key engineering concepts to grasp the gas turbine design principles and carry out nearly all calculations using simple formulas and a few technology charts. Furthermore, this will also enable the reader to critically evaluate new developments and claims to future improvements in electric power generation systems that employ turbomachines and utilize fossil fuels such as natural gas.

As Descartes wrote in his *Discourse on Method* “one cannot conceive anything so strange and so implausible that it has not already been said by one philosopher or another.” As such, in relation to the treatment of the subject in the paragraphs below, a claim of pure originality would be absurd. Nevertheless, the material is wholly original in the sense that key concepts are developed in a unique manner along with all pertinent formulas, representative design parameters and technology curves so that it forms a coherent and compact reference to be used in basic engineering analysis of gas turbine combined cycles. Even then, it must be pointed out that a similar treatment can be found in the excellent brief chapter on gas turbines written by Haselbacher cited in Ref. [26]. Another valuable short reference incorporating the basic fundamental considerations is cited in Ref. [24]. The treatment herein

focuses on land-based heavy-duty industrial gas turbines. For a brief introduction to thermodynamic and economic considerations pertaining to aeroderivative gas turbines and pertinent references, the reader should consult Horlock [27].

Basic Thermodynamics

Gas turbine is a relatively simple turbomachine comprising three key components: Compressor, combustor, and expander (Fig. 6). The last one is commonly referred to as a turbine. The compressor and expander are connected through a common shaft. The operation of a gas turbine is described by a thermodynamic cycle comprising four processes: compression, heat addition or combustion, expansion, and heat rejection. A gas turbine can also be classified as an internal combustion engine (just like a car engine), in which the compressed working fluid, i.e., air, mixes with fuel, and the products of the ensuing chemical reaction (i.e., combustion) expand through the turbine. In land-based electric power generation applications, part of the useful mechanical or shaft work, produced by the expanding combustion products (approximately 50% of the total) is consumed by the compressor. The remainder is converted into electric power in a synchronous alternating current (ac) generator, which is connected to the same shaft as the gas turbine.



Natural Gas Power. Figure 6

Modern heavy-duty industrial gas turbine (Courtesy: GE Energy)

In aircraft propulsion applications, the entire shaft work generated in the turbine is utilized to drive the engine compressor and fan. The thermal energy in the exhaust gas is converted into kinetic energy in the exhaust nozzle, which generates the net engine thrust.

The thermodynamic cycle that describes the gas turbine operation is the *Brayton cycle*. For simple analytical calculations, which can be found in elementary textbooks on thermodynamics, the ideal or *air-standard* Brayton cycle is used. The classical (and, from an engineering perspective, the most logical) representation of the air-standard Brayton cycle is on a temperature–entropy or T-s surface as illustrated in Fig. 7. The key assumptions in air-standard cycle analysis are:

1. The working fluid is air, which is a calorically perfect gas (i.e., $c_p = \text{constant}$).
2. Isentropic (i.e., adiabatic and reversible) compression and expansion.
3. Constant pressure heat addition and heat rejection.

The two key nondimensional performance metrics for a gas turbine are specific *net* power output and thermal efficiency:

$$w = \frac{\dot{W}_{\text{net}}}{\dot{m} \cdot c_p \cdot T_1} \quad (1)$$

$$\eta = \frac{\dot{W}_{\text{net}}}{\dot{Q}_{\text{in}}} \quad (2)$$

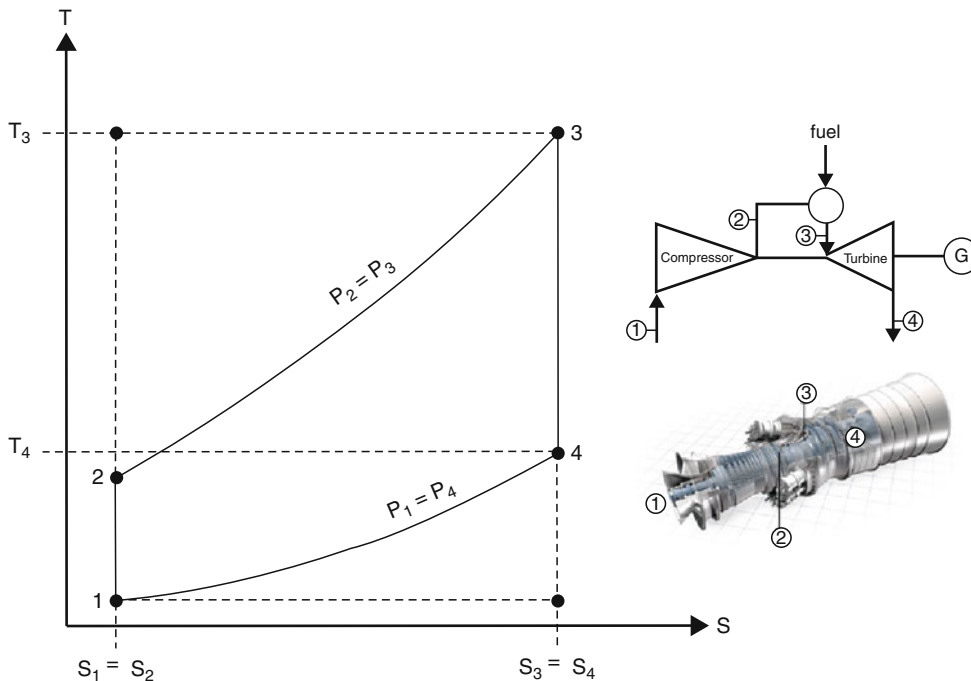
The net power output is the difference between the compressor power consumption and the expander power production. In graphical terms, it is exactly equal to the area encompassed by the cycle 1-2-3-4-1 on the T-s diagram in Fig. 7. Each component's power can be calculated by the application of the *first law of thermodynamics* for steady-state steady-flow (SSSF) successively to their respective *control volumes*:

$$\dot{W}_{\text{comp}} = \dot{m} \cdot c_p \cdot (T_1 - T_2) \quad (3)$$

$$\dot{W}_{\text{turb}} = \dot{m} \cdot c_p \cdot (T_3 - T_4) \quad (4)$$

$$\dot{W}_{\text{net}} = \dot{W}_{\text{turb}} + \dot{W}_{\text{comp}} \quad (5)$$

From Fig. 7 and (Eqs. 3 and 4) one can easily see that the compressor power consumption is a *negative* number (i.e., work done *on* the control volume) and expander power generation is a *positive* number



Natural Gas Power. Figure 7
Temperature–entropy diagram of gas turbine Brayton cycle

(i.e., work done *by* the control volume). Application of the first law for SSSF to the heat addition and heat rejection processes, one gets

$$\dot{Q}_{\text{in}} = \dot{m} \cdot c_p \cdot (T_3 - T_2) \quad (6)$$

$$\dot{Q}_{\text{out}} = \dot{m} \cdot c_p \cdot (T_1 - T_4) \quad (7)$$

From Fig. 7 and (Eqs. 6 and 7), the heat input to the cycle is a *positive* number (i.e., heat supplied *to* the control volume) and heat rejection from the cycle is a *negative* number (i.e., heat taken *from* the control volume).

In addition to the first law of thermodynamics for a SSSF process (or control volume analysis), there is one more key thermodynamic relationship that forms the heart of the GT Brayton cycle calculations: pressure–temperature ratio for an isentropic process:

$$\frac{T_2}{T_1} = \left(\frac{p_2}{p_1} \right)^{\frac{\gamma-1}{\gamma}} \quad (8)$$

where γ is the ratio of constant pressure and constant volume-specific heats. Combining (Eqs. 1–7), in nondimensional terms and applying the isentropic formula in Eq. (8) to compression and expansion processes, one ends up with the following two expressions:

$$w = (\pi^k - 1) \cdot \left(\frac{\tau_3}{\pi^k} - 1 \right) \quad (9)$$

$$\eta = 1 - \frac{1}{\pi^k} \quad (10)$$

where $\pi = p_2/p_1$ is the cycle pressure ratio, τ_3 is the cycle maximum temperature (nondimensionalized via division by T_1) and k is the isentropic exponent, $\gamma - 1/\gamma$. The two simple formulas, Eqs. (9) and (10), which are strictly valid only for an ideal air-standard cycle, illustrate almost all of the key facts concerning gas turbine performance:

1. Gas turbine or Brayton cycle performance is controlled by two parameters: Cycle pressure ratio, π , and maximum cycle temperature, τ_3 .
2. Cycle efficiency is a function of cycle pressure ratio only; higher cycle pressure leads to higher cycle efficiency.

3. Higher cycle temperature leads to higher cycle specific output.
4. Higher cycle pressure ratio, beyond a certain value, is detrimental to cycle specific output (i.e., $\frac{\partial^2 w}{\partial \pi^2} < 0$ and $\pi = \sqrt[k]{\tau_3}$ for $\partial w / \partial \pi = 0$).

These observations are also borne out qualitatively by a close examination of the cycle T-s diagram in Fig. 7. As the heat addition isobar moves upward (i.e., higher cycle pressure ratio, π) one can visually verify that the area encompassed by the cycle, 1-2-3-4-1, becomes smaller until, in the limit, it becomes zero. Equation (10) numerically implies that, in the limit when π is sufficiently high, the cycle efficiency will approach 100%, or, $\eta \rightarrow 100\%$ as $\pi \rightarrow \infty$. This, of course, is impossible, because, as dictated by the *second law of thermodynamics*, the maximum theoretical efficiency attainable by any cycle is given by the Carnot cycle efficiency. The Carnot cycle is a hypothetical power cycle that is represented by two isentropic processes, i.e., compression and expansion; and two isothermal processes, i.e., heat addition and heat rejection. Graphically, it is the rectangle enveloping the Brayton cycle, which is bounded by T_1 and T_3 isotherms and s_1 and s_3 isentropes.

For the cycle in Fig. 7, one can easily visualize the limit of $\pi \rightarrow \infty$ as T_3 stays constant. In the limiting case, one will end up with a Brayton cycle that is a rectangle with ~ 0 area (i.e., zero specific output) and heat addition and rejection at constant temperatures T_3 and T_1 , respectively, which essentially defines a Carnot cycle with the efficiency

$$\eta = 1 - \frac{T_1}{T_3} \quad (11)$$

Comparing with Eq. (10), for a cycle maximum temperature of $T_3 = 1,500^\circ\text{C}$ and $\gamma = 1.4$, this would correspond to a cycle pressure ratio of $\pi = 577$ and cycle efficiency $\eta = 87.4\%$. In other words, as dictated by the second law of thermodynamics, the maximum theoretical cycle efficiency that is attainable by a gas turbine with a “firing temperature” of $1,500^\circ\text{C}$ is 87.4%. Clearly a cycle pressure ratio of 577 is of little practical interest. For the modern heavy-duty industrial gas turbines, the typical value of π is about 18. What does the second law say about a gas turbine for a cycle maximum temperature of $T_3 = 1,500^\circ\text{C}$ and $\pi = 18$? In order to

find that out, consider the isobaric heat addition process $2 \rightarrow 3$ in Fig. 7. For a reversible and isobaric process, the modified Gibbs equation is

$$h_3 - h_2 = \int_2^3 T \cdot ds \quad (12)$$

One can imagine an (hypothetical) isothermal process that takes place at a constant temperature, whose value lies between T_3 and T_2 , and has the same entropy as the actual end states 2 and 3. This process would have exactly the same numerical value of heat addition as the original process, i.e., $h_3 - h_2$, per unit mass. For that process, Eq. (12) can be rewritten as

$$h_3 - h_2 = \bar{T}_H \cdot (s_3 - s_2) \quad (13)$$

Using the calorically perfect gas relationships, the mean-effective heat addition temperature can be found from Eq. (13) as

$$\bar{T}_H = \frac{T_3 - T_2}{\ln\left(\frac{T_3}{T_2}\right)} \quad (14)$$

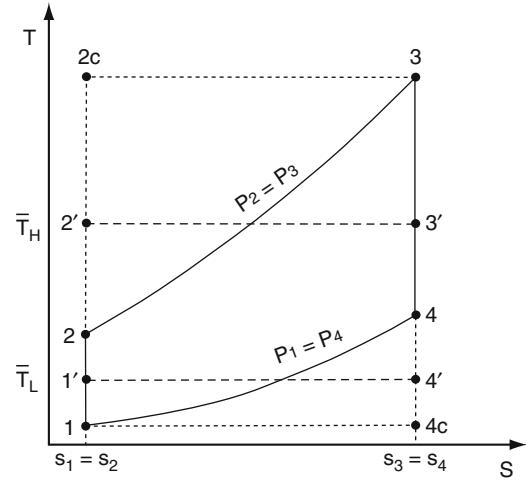
Using the same logic, the mean-effective heat rejection temperature is calculated as

$$\bar{T}_L = \frac{T_4 - T_1}{\ln\left(\frac{T_4}{T_1}\right)} \quad (15)$$

These two mean-effective temperatures are graphically depicted in Fig. 8. Thus, qualitatively any thermodynamic cycle in general, and the Brayton cycle in particular, 1-2-3-4 in Fig. 8, can be represented by its *Carnot-equivalent* cycle, 1'-2'-3'-4', with mean-effective (constant) heat addition and heat rejection temperatures that are given by Eqs. (14 and 15). The efficiency of that cycle would be given by the well-known relationship for the Carnot efficiency, i.e.,

$$\eta = 1 - \frac{\bar{T}_L}{\bar{T}_H} \quad (16)$$

For an ideal air-standard Brayton cycle, one can show that Eq. (16) is the same as Eq. (10). Thus, for any gas turbine with a cycle pressure ratio of 18, the maximum theoretical efficiency is 56.2%. Using Eqs. (8), (14), and (15), this would be the efficiency of a Carnot cycle with isothermal heat addition and heat rejection at 852°C and 219°C, respectively. It is



Natural Gas Power. Figure 8

Temperature–entropy diagrams of gas turbine Brayton cycle and its “Carnot equivalent”

interesting to note that, changing the cycle maximum temperature from 1,500°C to 1,400°C would change these temperatures to 814°C and 203°C, respectively, but would *not* change the cycle efficiency, i.e., 56.2%, which is only a function of π . This is in contrast to the performance trend of the “real” gas turbines, whose thermal efficiency does indeed increase with increasing cycle maximum or *firing* temperature. Why this is so will be understood below.

One key observation from Fig. 8 is the large thermal efficiency gap between the ultimate Carnot efficiency given by Eq. (11) and the maximum theoretical entitlement level as dictated by the second law, Eq. (10) or Eq. (16), which is nearly 30 percentage points. Graphically, this gap is represented by the two triangular areas in Fig. 8:

1. The upper triangular area 2-2c-3-2', which is equal to the rectangular area 2'-2c-3-3'-2', and represents the heat addition (i.e., combustion) irreversibility
2. The lower triangular area 1-1'-4'-4c-1, which is equal to the rectangular area 1-1'-4'-4c-1, and represents the irreversibility and exergy loss via heat rejection to a heat sink

In terms of cycle temperatures, the upper triangular area is quantified by the gap between the cycle maximum temperature, T_3 , and the mean-effective heat addition temperature \bar{T}_H . Similarly, the lower

triangular area is quantified by the gap between the cycle minimum or ambient temperature, T_1 , and the mean-effective heat rejection temperature \bar{T}_L . The latter mechanism of “lost work” associated with heat rejection also points the way to the gas and steam turbine combined cycle power plant concept. This will be explored in some more detail later in the discussion.

The real gas turbine Brayton cycle differs from the ideal, air-standard cycle in the following respects:

1. Non-isentropic compression and expansion (represented by component polytropic or isentropic efficiencies)
2. Pressure losses during heat addition and heat rejection
3. Real gas effects (represented by a suitable equation of state)
4. Change in working fluid composition and properties downstream of combustion process

These nonidealities can be easily incorporated into the simple equations above via polytropic component efficiencies, pressure loss factors, and real gas equation of state. One can refer to one of the available textbooks on the subject for details, e.g., Ref. [29]. For example, Eq. (9) can be rewritten as

$$w = \left(1 - \pi_c^{\frac{k_a}{\eta_c}}\right) + (1 + f) \cdot \chi_g \cdot \tau_3 \cdot \left(1 - \frac{1}{\pi_t^{\frac{\eta_t}{\eta_c \cdot k_g}}}\right) \quad (17)$$

where π_c and $\pi_t < \pi_c$ are compressor and turbine pressure ratios, respectively, k_a and k_g are isentropic exponents for air and combustion products, respectively, and η_c and η_t are compressor and turbine polytropic efficiencies, respectively. The ratio of the specific heats of combustion products and air is χ_g and f is the fuel–air mass flow ratio. Heat addition can be found from the fuel consumption, i.e.,

$$\dot{Q}_{in} = \dot{m}_f \cdot \text{LHV} \quad (18)$$

where LHV is the net or lower heating value of the fuel at the reference temperature, T_{ref} . The efficiency of the gas turbine is the ratio of Eqs. (17) and (18), i.e.,

$$\eta = \frac{\left(1 - \pi_c^{\frac{k_a}{\eta_c}}\right) + (1 + f) \cdot \chi_g \cdot \tau_3 \cdot \left(1 - \frac{1}{\pi_t^{\frac{\eta_t}{\eta_c \cdot k_g}}}\right)}{f \cdot \ell} \quad (19)$$

where ℓ is the nondimensional fuel energy content, $\text{LHV}/c_{p,a} \cdot T_1$. Note that for clarity, mechanical and electric losses are omitted from Eqs. (17) and (19), which represent the “shaft” performance. Typically quantified via mechanical and generator efficiencies, η_m and η_g , respectively, these losses should be accounted for the gas turbine generator’s net electric output. The fuel mass flow rate (as a fraction of turbine airflow) that is requisite for a specified cycle maximum temperature is given by

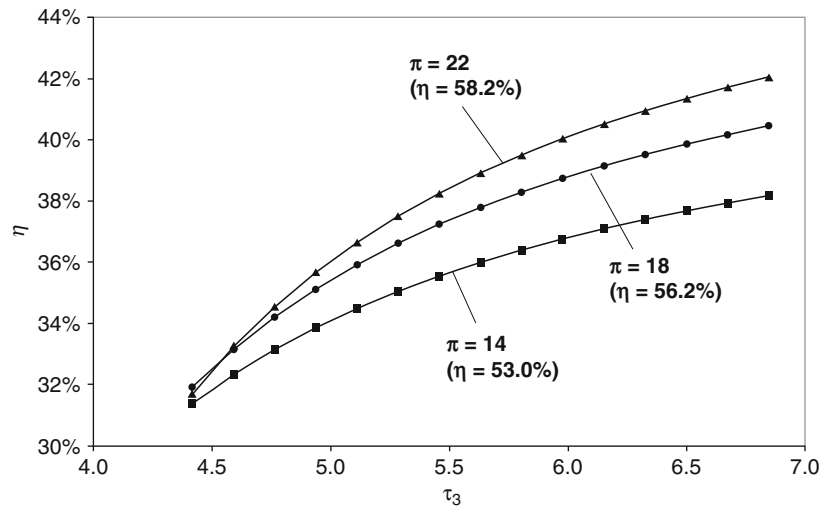
$$f = \frac{\chi_g \cdot (T_3 - T_{ref}) - (T_2 - T_{ref})}{\ell \cdot T_1 + \chi_f \cdot (T_f - T_{ref}) - \chi_g \cdot (T_3 - T_{ref})} \quad (20)$$

where the second term in the denominator is the sensible fuel energy input and χ_f is the ratio of the specific heats of fuel and air. Comparing Eq. (19) (along with Eq. (20) that sets the denominator) with Eq. (10), one can easily observe that the GT thermal efficiency, when accounting for the “real” cycle effects, is indeed a function of τ_3 . While it is difficult to discern the impact of τ_3 on η just by looking at Eqs. (19) and (20), a sample calculation can illustrate that as shown in Fig. 9.

Typical values that are reasonably accurate for gas turbines burning natural gas fuel (assumed to be 100% CH₄ or methane) are given below (adopted from Ref. [29]) and used in the calculations that resulted in the curves depicted in Fig. 9:

$$\begin{aligned} \text{LHV} &= 21,515 \text{ Btu/lb} \\ c_{p,a} &= 0.24 \text{ Btu/lb-R (1.005 kJ/kg-K)} \\ c_{p,g} &= 0.274 \text{ Btu/lb-R (1.148 kJ/kg-K) or } \chi_g = 1.1417 \\ c_{p,f} &= 0.60 \text{ Btu/lb-R (2.5125 kJ/kg-K) or } \chi_f = 2.50 \\ \gamma_a &= 1.400 \\ \gamma_g &= 1.333 \\ \eta_c &= 87.8\% \\ \eta_t &= 85\% \\ \eta_m &= 99\% \\ \eta_g &= 98.9\% \\ k_a &= 0.2857 \\ k_g &= 0.2498 \\ T_{ref} &= 77^\circ\text{F (25}^\circ\text{C)} \\ \ell &= 172.728 \end{aligned}$$

Original equipment manufacturer (OEM) data for five 50-Hz (3,000 rpm) heavy-duty industrial gas



Natural Gas Power. Figure 9
 Real Brayton cycle efficiency from Eqs. (19) and (20) as a function of cycle pressure ratio (π) and maximum cycle temperature (τ_3). Also shown (in parentheses) are the ideal, air-standard cycle efficiencies from Eq. (10)

Natural Gas Power. Table 2 OEM data for 50-Hz heavy-duty industrial gas turbines [6]. The maximum cycle temperature, T_3 , values are estimated using the published output, efficiency, pressure ratio (PR), and exhaust data with Eqs. (17–20)

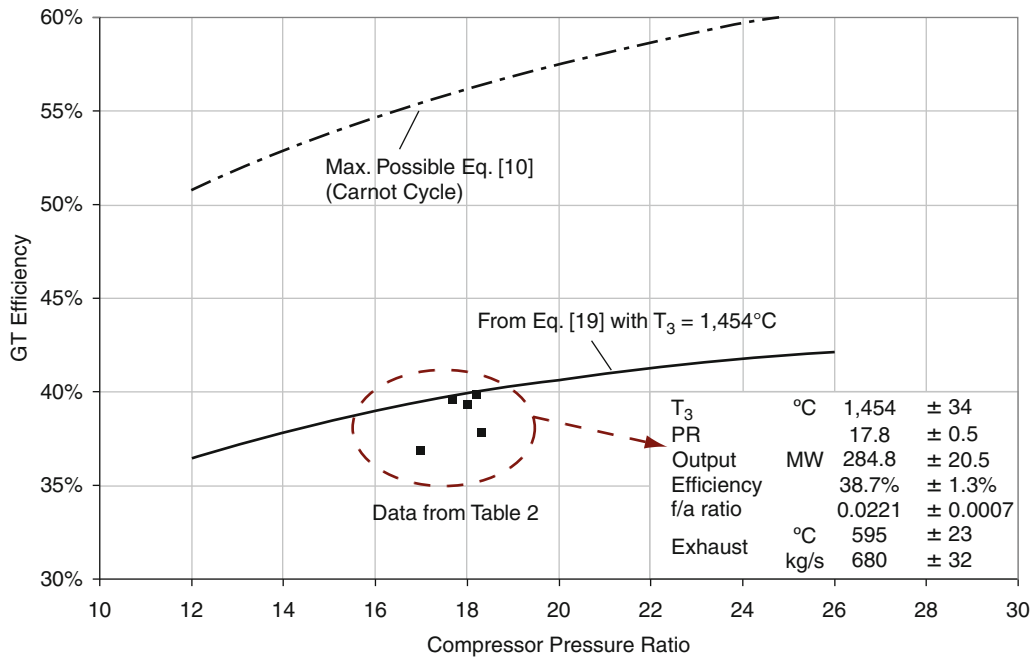
Turbine	T_3						Exhaust	
	(°C)	°F	PR	PR _t	Output (MW)	Eff.	(kg/s)	°C
A	1,415	2,579	17.7	15.9	285.0	39.6%	690	572
B	1,443	2,629	17.0	13.5	255.6	36.9%	641	602
C	1,501	2,734	18.3	14.5	279.2	37.8%	655	629
D	1,435	2,616	18.2	16.3	292.0	39.8%	692	577
E	1,476	2,688	18.0	15.5	312.1	39.3%	720	597

turbines from a recent trade publication [6] are given in Table 2.

For an actual gas turbine, the cycle maximum temperature T_3 is equivalent to the turbine inlet temperature (TIT) at the exit of the gas turbine combustor. Another value that is frequently used in the industry is the “firing temperature,” which is the temperature at the stage-1 blade row (rotor) inlet. This temperature, also known as rotor inlet temperature (RIT) is the maximum temperature at which the combustion products start producing useful expansion work. The difference between TIT and RIT is about 250°F for modern air-cooled gas turbines and reflects the

dilution effect of the stage-1 nozzle and forward wheel space cooling flow. One other cycle maximum temperature definition, favored by European OEMs, is the ISO-TIT, which is a fictitious number as defined by ISO-2314 (typically ~180°F lower than RIT). Obviously, for an uncooled turbine all these definitions are the same.

Gas turbine efficiency as a function of PR is shown in Fig. 10. Specific output, w , in the range covered by the turbines in Table 2 is 400–450 kJ/kg. Carbon dioxide emissions are calculated from the airflow and fuel flow using chemical reaction equations for 100% CH₄ gaseous fuel assumption as 510 kg/h per MW of output



Natural Gas Power. Figure 10

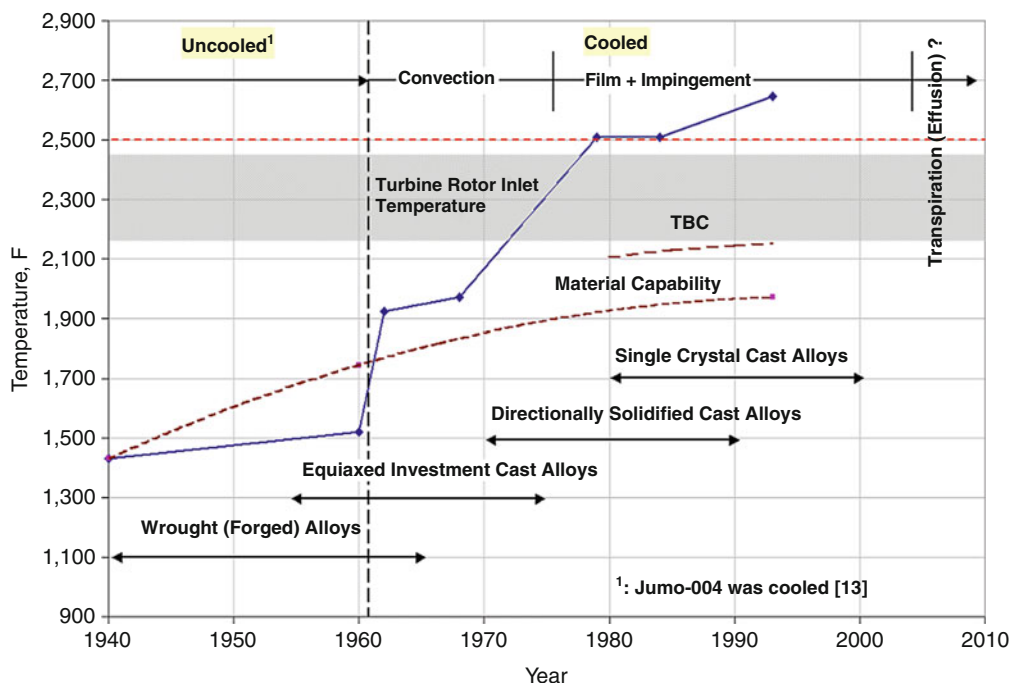
Gas turbine efficiency from Eq. (19) as a function of compressor PR (Data is from Ref. [6] and summarized in Table 2)

at 39% net efficiency. Each one percentage point improvement in efficiency is worth about 15 kg/h per MW reduction in CO₂ emissions. This is equivalent to 27,000 metric tons for a 300 MW GT running at base load for 6,000 h/year.

Figure 10 shows that the gap between the current State of the Art (SOA) in gas turbine technology and the maximum theoretical entitlement level, as dictated by the second law, is more than 15 efficiency points. The GT efficiency – compressor pressure ratio (PR) trend in Fig. 10 (via Eq. (19)) gives the impression that, for a given cycle maximum temperature, the efficiency can be improved indefinitely by increasing the PR. This is *not* true due to a key feature of modern gas turbines that is very difficult to capture in simplified models such as Eqs. (17–20): turbine hot gas path (HGP) component cooling via air extracted from the compressor discharge and/or interstage locations. Since the early days of GT development, the components exposed to the highest temperature environment have been cooled with air drawn from the compressor. The Junkers Jumo-004 jet engine of the German fighter plane Messerschmitt 262, which entered active service in 1944 during the last stages of the World War II, was

the first mass production turbojet engine in the history with a turbine inlet temperature (TIT) of 1,427°F [13]. The hollow turbine blades (manufactured by folding flat sheets of 12% Cr alloy called *Cromadur*) were cooled by air extracted from between the fourth and fifth stages of the compressor and introduced to the blade through the holes drilled into the disk. The cooling technology, although unchanged in basic principles, has advanced greatly over the years with the introduction of complex serpentine cooling passages, impingement, film, and transpiration cooling techniques; paced by concurrent advances in superalloys (e.g., nickel-based directionally solidified or single-crystal), manufacturing techniques (e.g., investment casting), oxidation/corrosion (e.g., MCrAlY), and thermal barrier (e.g., yttria-stabilized zirconia, YSZ) coatings. This development supported the advances in TIT, which nearly doubled the standard set by Jumo-004 over the course of the next 50 years (e.g., see Fig. 11 that uses data from Cumpsty [30]). A good overview of the current SOA in advanced GT materials and coatings can be found in Ref. [31].

An alternative to open loop air-cooling (OL-AC) has recently been commercially introduced at the



Natural Gas Power. Figure 11

Turbine RIT and bucket alloy capability history [30]. One can add $\sim 250^\circ\text{F}$ to RIT to get a rough idea about TIT of the advanced F-Class units. The shaded region designates the melting point of nickel-based superalloys

480-MW Baglan Bay 109H CC power plant in Wales [32]. This machine employs closed-loop steam cooling (CL-SC) for the majority of the turbine HGP components (the first two stages of the four-stage turbine, to be exact). A “lighter” version of the H-System™ is the G-Class turbine with steam cooling limited to the stationary parts upstream of the turbine HGP [33].

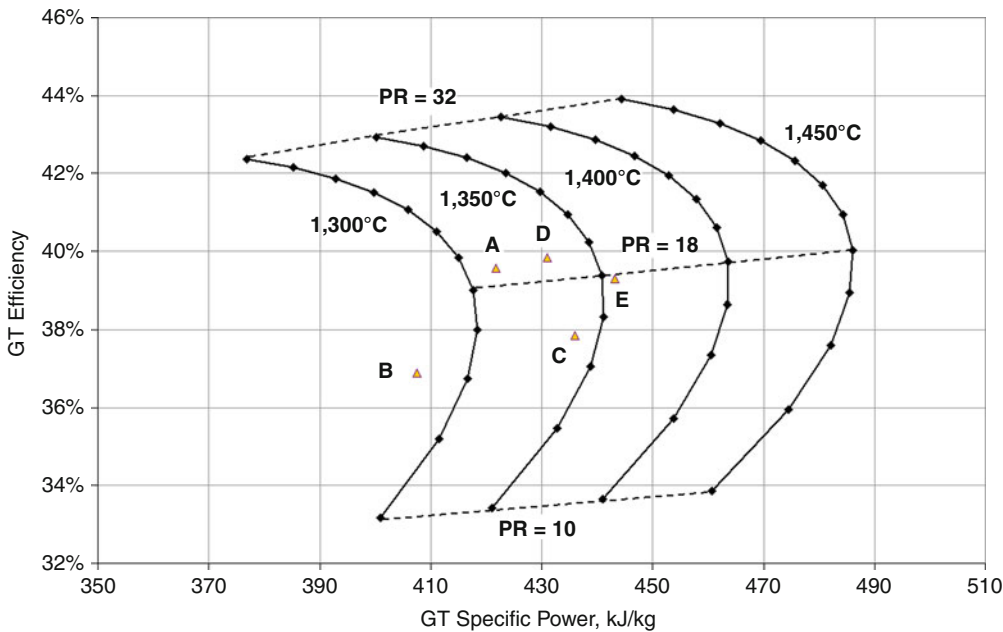
A structural refinement to the basic GT Brayton cycle is the inclusion of a second “reheat” combustor within the turbine expansion. Reheat, or *sequential combustion* as it is referred to by the OEM that successfully commercialized the technology increases the mean-effective cycle heat addition temperature and reduces the combustion irreversibility [34] (the upper triangular area 2-2c-3-2 in Fig. 3). From a fundamental thermodynamic perspective, lower combustion losses give the GT cycle with reheat combustion a higher performance entitlement than a cycle without it. Similarly, there is no doubt that closed-loop (external) cooling is superior to open-loop (internal) cooling but the exact magnitude of that advantage in real machines is difficult to quantify. Chiesa and Macchi

[35] quantified the advantages of OL-AC (with and without reheat combustion) and CL-SC technologies via detailed calculations. Using OEM data from GE’s 7FB, Siemens V94.3A (now SGT5-4000F), Alstom’s GT-26, and MHI’s M701G units for calibration, they have developed very detailed cooled GT models and evaluated the CC performances. Their findings are summarized in Table 3, which shows that CL-SC has an advantage of ~ 1.8 points (%) in net η_{CC} over OL-AC.

A realistic assessment of the gas turbine requires some means to incorporate the cooling penalty into the GT calculations. Cooled turbine expansion models, no matter how much simplified, involve a rather cumbersome system of equations amenable only to a numerical solution. In the open literature, there are numerous published works that range from Holland and Thake [36] and several papers by Elmasri [37–41] during 1980s to the more recent works of Horlock and coworkers [42–45]. Young and Wilcock’s comprehensive study [45] and the similar treatment in Ref. [35] incorporate detailed aerothermodynamic calculations.

Natural Gas Power. Table 3 Comparison of CC performances based on four major GT technologies [35]: Air-cooled (OL-AC), air-cooled reheat (OL-AC+SC), steam-cooled (CL-SC), and steam-cooled reheat (CL-SC+SC)

	OL-AC	OL-AC+SC	CL-SC	CL-SC+SC
GT PR	20	30	23	30
HPT RIT (°C)	N/A	1,290	N/A	1,430
HPT RIT (°F)	N/A	2,354	N/A	2,606
RIT (°C)	1,430	1,430	1,430	1,430
RIT (°F)	2,606	2,606	2,606	2,606
Net η_{CC} (% , points)	Base	+0.81	+1.77	+2.74



Natural Gas Power. Figure 12

“Real” gas turbine technology curves for different PR-RIT combinations [46] (Also shown is data from Table 2 (as published in Ref. [6]))

On the business side, obviously not available for public consumption, OEMs have their own in-house proprietary codes containing the field-proven knowledge (instead of theoretical approximations) distilled from decades of experience in manufacturing and testing gas turbines. The results from a more realistic GT calculation [46], which includes the impact of turbine cooling with compressor extraction air, are shown in Fig. 12.

Myriad alternate GT design variations have been proposed and entered the literature, some to the extent

that they are even covered in elementary textbooks; e.g., compressor intercooling, regeneration (also referred to as recuperation), a combination of both, steam injection (STIG) [47, 48], humid air turbine (HAT) [49], chemical recuperation [50], and closed cycles [51] are among the most well known. For a comprehensive look at those variants and their relative merits, as well as references to the pertinent literature, one can consult the relevant sections in Wilson and Korakianitis [11]. Intercooling and recuperation

technologies have been demonstrated successfully in aeroderivative gas turbines [52,53], whereas the combination of both (ICR) has been incorporated into a ship-propulsion unit [54]. Nevertheless, steam-cooled (G- or H-Class) and reheat machines are the only two commercially proven and successful heavy-duty industrial variants of the basic air-cooled GT, which are suitable to large-scale base load electric power generation.

So far it has been demonstrated that increasing the compressor PR and/or TIT (i.e., maximum cycle temperature) increases the GT efficiency and specific work. The latter reaches a maximum before further increase in PR decreases it. Furthermore, increasing TIT is favorable to CC-specific work and CC efficiency. These trends can be easily visualized and quantified by the ideal cycle diagrams shown in Fig. 13.

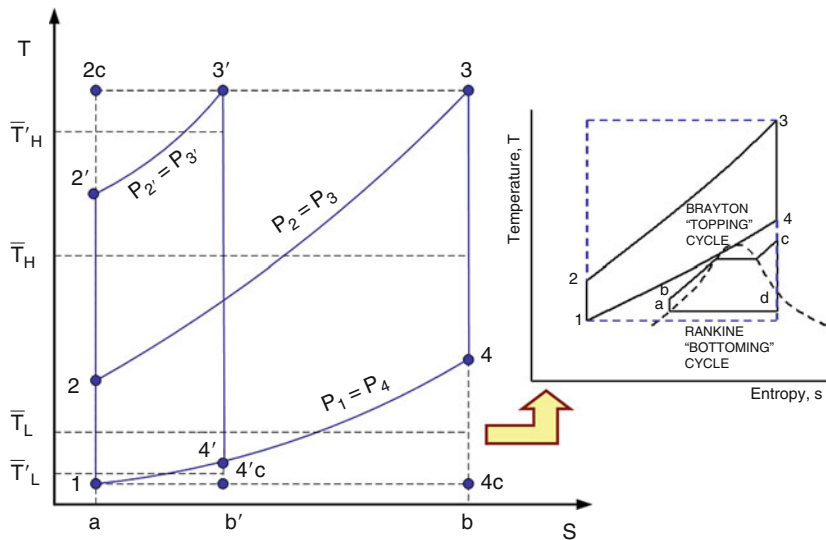
The heat rejection “lost work” quantified by the lower triangular area 1-4-4c-1 suggests that a better approximation of the Carnot cycle can be achieved by utilizing the heat source provided by the working fluid at the GT exhaust (i.e., state 4) in a second power generation cycle. At this point, the Rankine cycle presents itself as the ideal choice due to two facts:

1. Good match with GT exhaust (on average about 600°C for SOA F-Class machines) for steam generation at 125–165 bar

2. Constant temperature heat rejection (i.e., in the steam condenser) at a temperature that is much closer to T_1 than \bar{T}_L

Earlier in the section, it was shown that the Brayton cycle efficiency is given by the mean-effective cycle heat addition and rejection temperatures, \bar{T}_H and \bar{T}_L , respectively, e.g., Eq. (16). Increasing the cycle maximum pressure from p_2 to p_2' increases the cycle efficiency by simultaneously increasing \bar{T}_H and reducing \bar{T}_L . It is easy to see that the cycle-specific work, quantified by the area 1-2-3-4, would go through a maximum before approaching zero at which point the cycle efficiency is the Carnot efficiency. The ultimately detrimental effect of continuously increasing the Brayton “topping” cycle (BTC) PR on CC efficiency can also be seen from Fig. 13. The maximum possible CC “bottoming” cycle work, limited by the second law, is equal to the triangular area 1-4-4c-1, which can be shown to be exactly equal to the GT exhaust gas exergy [55]. It is easy to visually appreciate the steady reduction in this maximum with increasing PR (cf. 1-4-4c-1 and 1-4'-4'c-1). The efficiency of the ideal Rankine bottoming cycle (RBC) is given by the equation below:

$$\eta_{\text{RBC}} = 1 - \frac{T_1}{T_L} \quad (21)$$



Natural Gas Power. Figure 13

Comparison of Brayton cycles with different PRs on a T-s diagram. Also illustrated is the “combined” cycle concept, in which the Rankine “bottoming” cycle partially recovers the Brayton “topping” cycle heat rejection “lost” work

where \bar{T}_L is from Eq. (15). The formula in Eq. (21) clearly implies that this particular ideal Rankine cycle would be equivalent to a Carnot engine operating between the temperature reservoirs at \bar{T}_L and T_1 . Furthermore, the product of Eq. (21) and the total energy content of the GT exhaust \dot{Q}_{exh} can be shown to be exactly equal to the total exergy of the exhaust gas [55]. Exergy is a fluid property that can be calculated using a suitable equation of state with two other known properties (e.g., pressure and temperature) and the composition.

By looking at the cycle T-s diagram in Fig. 13, the ideal CC efficiency with ideal Brayton and Rankine cycles can be deduced to depend on the ratio of the mean-effective Brayton cycle heat addition temperature given by Eq. (14) and the ambient temperature T_1 , i.e.,

$$\eta_{\text{CC}} = 1 - \frac{\ln\left(\frac{\tau_3}{\pi^k}\right)}{\tau_3 - \pi^k} \quad (22)$$

Ignoring the miscellaneous topping and bottoming cycle losses and minor inputs, a simplified version for the CC shaft efficiency can also be written as

$$\eta_{\text{CC}} = \eta_{\text{BTC}} + (1 - \eta_{\text{BTC}}) \cdot \eta_{\text{RBC}} \quad (23)$$

Taking the derivative of both sides with respect to the PR and setting the CC efficiency derivative to zero for the maximum, one finds that (note that the RBC and BTC efficiencies are approximately the same in magnitude)

$$\frac{\partial \eta_{\text{BTC}}}{\partial \text{PR}} \approx - \frac{\partial \eta_{\text{RBC}}}{\partial \text{PR}} \quad (24)$$

This relationship states that the maximum *Brayton–Rankine* combined cycle efficiency occurs at the point where the rate of increase of BTC efficiency with PR is the same as the rate of decrease of the RBC efficiency. Using the ideal Brayton cycle formulas, which can be found in any standard thermodynamics textbook such as Ref. [28], the rate of increase of BTC efficiency with increasing PR is

$$\frac{\partial \eta_{\text{BTC}}}{\partial \text{PR}} = \frac{\gamma - 1}{\gamma} \cdot \text{PR}^{\frac{1-2\gamma}{\gamma}} \quad (25)$$

From Eq. (21), once again using the ideal Brayton cycle relationships, the rate of decrease of the RBC efficiency with increasing PR is found as

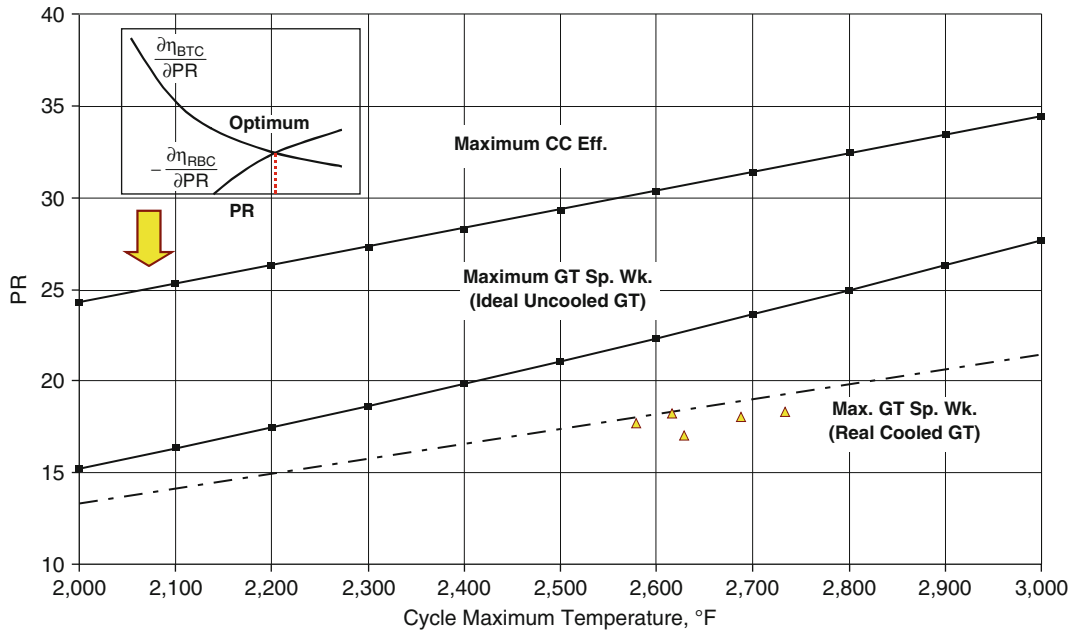
$$\frac{\partial \eta_{\text{RBC}}}{\partial \text{PR}} = \frac{\gamma - 1}{\gamma} \cdot \frac{\left(\frac{T_1}{\bar{T}_L}\right) \cdot \left(\frac{T_4}{T_1} - 1\right)}{\text{PR} \cdot \ln\left(\frac{T_4}{T_1}\right)} \quad (26)$$

Combining Eqs. (21–26) with $\gamma = 1.4$ and $T_1 = 85^\circ\text{F}$ (assuming RBC heat rejection at 1.2 in. Hg or 41 mbar of steam condenser pressure) and solving them numerically, the PR of the BTC for maximum CC efficiency can be plotted as a function of T_3 (see Fig. 14). Also plotted in Fig. 14 is the PR corresponding to the maximum BTC specific work, $\pi = \sqrt[3]{\tau_3}$, with $T_1 = 59^\circ\text{F}$. For comparison, representative OEM data extracted from a recent trade publication [6] is also shown. It is interesting to note the excellent agreement of the PR-TIT data of actual production machines with values corresponding to maximum GT-specific power output predicted using a rigorous cooled gas turbine model.

Using an uncooled *quasi-ideal* GT model with reasonable assumptions outlined earlier in the section to take into account the real turbine losses, CC efficiencies corresponding to the parameters in Fig. 14 are calculated and plotted in Fig. 15. The key takeaways from the plots in those figures are rather illuminating:

1. Actual OEM designs are at PR values somewhat lower than those for maximum GT-specific output in an uncooled GT but coincide extremely well with the air-cooled GT specific work line, in accordance with the earlier observation made by Horlock [24].
2. This seems to be a reasonable choice from a cost perspective with ~ 0.50 percentage points sacrifice in performance vis-à-vis the maximum η_{CC} that has a $\sim 50\%$ larger PR. In fact, the proximity of the two maxima is even closer in the presence of turbine cooling so that this is a very reasonable trade-off indeed.
3. 65% and 70% net CC efficiencies require uncooled turbines with TIT of $2,500^\circ\text{F}$ and $3,000^\circ\text{F}$, respectively.

OEM data for combined cycle performances based on five 50-Hz (3,000 rpm) heavy-duty industrial gas turbines from a recent trade publication [6] (see Table 2) are given in Table 4. The data is for two major combined cycle configurations, which are labeled as 1×1 (i.e., one GT and one ST) and 2×1 (i.e., two GTs and one ST). Using the GT performance



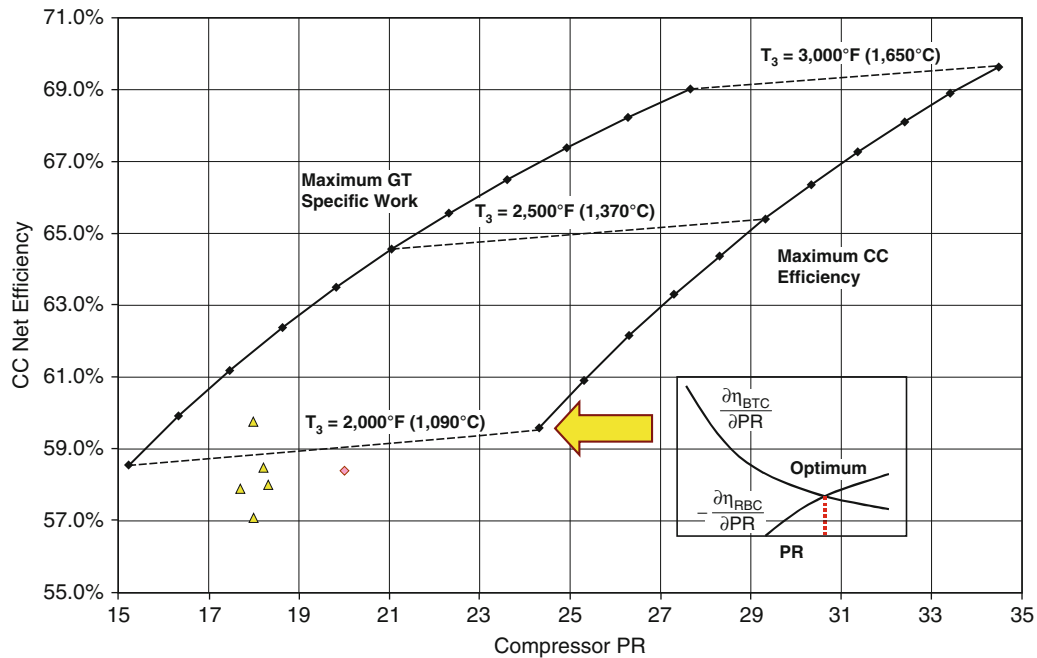
Natural Gas Power. Figure 14

PR for maximum CC efficiency and maximum BTC specific work. Data points are from Ref. [6]. Cycle maximum temperature or T_3 values are from Table 2. The inset illustrates the calculation of maximum η_{CC} point using Eqs. (24–26)

in Fig. 12 as a basis, CC performance curves can be developed using the exergy-based bottoming cycle calculations in Ref. [56] (see Fig. 16).

Today's SOA in CC performance with advanced gas turbines corresponds to a range of 57% to ~60% with TIT values in the range of 2,600–2,800°F. This is a far cry from the theoretically possible values in the absence of turbine cooling as shown in Fig. 15 (by about 10 percentage points). Sponsored by the DOE's National Energy Technology Laboratory (NETL), the *Vision 21* program has a goal to develop, by 2015, the core modules for a fleet of fuel-flexible, multiproduct energy plants that boost power efficiencies to more than 60% (75% on an LHV basis for gas-fueled plants), emit virtually no pollutants, and with carbon sequestration release minimal or no carbon emissions. Lately, several published claims about the future of the CC plant performance at the end of the first quarter of the twenty-first century stated a goal of 65% with the possibility of over 70% with an air-cooled GT [57]. The uncooled GT-based CC efficiency curves and the current SOA data in Fig. 15 underline the significant hurdles to be overcome in terms of cooling flow reduction

in pace with increasing firing temperatures in order to achieve those goals. The gap between the uncooled GT TIT (i.e., same as RIT) requirement for 65% η_{CC} and the current material capability is nearly 800°F (~450°C). The current trend in superalloy capabilities, considering the competing tendencies in strength (i.e., creep rupture and fatigue limits) and corrosion resistance, does not seem to be likely to alleviate too much of that [58]. A combination of advanced materials including ceramic matrix composites, metal foams (to enable the upper limit of film cooling, i.e., effusion or transpiration cooling) [59] and external cooling (most likely via steam) accompanied by bottoming cycle advances is imperative to even approach the stated efficiency goals with acceptable NO_x and CO emissions. It is thus fairly obvious that a η_{CC} goal of 65% and higher in a *Brayton–Rankine* CC power plant is a highly unlikely proposition in near future. This is more or less the conclusion reached by Rao et al. [60] who investigated the plant system configurations that might be able to meet the performance and emission goals of *Vision 21* program. They concluded that hybrid systems integrating fuel cells with gas turbines,



Natural Gas Power. Figure 15

Uncooled GT-based CC efficiencies (note that $T_3 \equiv \text{TIT}$). The data (symbol Δ) is from Table 4. The other data is from Ref. [35] for the SOA air-cooled GT (TIT=2,826°F). The inset illustrates the calculation of maximum η_{CC} point using Eqs. (24)–(26) in the text. Component polytropic efficiencies are assumed to be 92.5% and 90% for the compressor and turbine, respectively

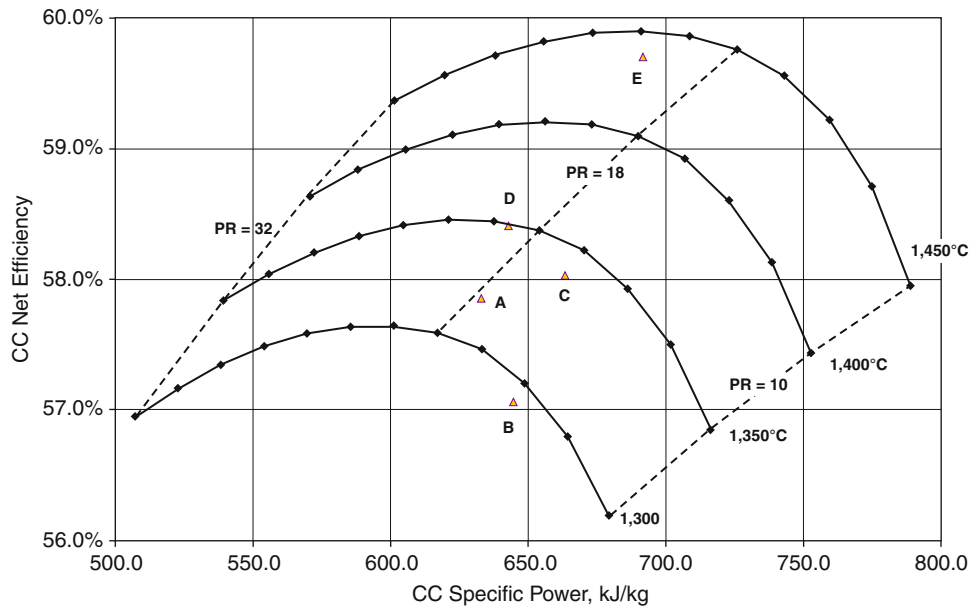
N

Natural Gas Power. Table 4 OEM data for combined cycle performances (1×1 and 2×1) of 50-Hz heavy-duty industrial gas turbines [6] (Gas turbine data is in Table 2)

Turbine	GT MW	1×1			2×1		
		ST	CC		ST	CC	
		MW	MW	η (%)	MW	MW	η (%)
A	285.0	141.0	415.1	57.9	282.0	830.2	57.9
B	255.6	141.8	390.8	56.7	289.2	786.9	57.1
C	279.2	157.9	412.9	58.0	315.3	825.4	58.0
D	292.0	144.7	423.0	58.4	290.3	848.0	58.5
E	312.1	157.3	459.0	59.5	317.7	921.1	59.7

e.g., SOFC/GT hybrid system [61], were required to meet the program goals. There's little doubt about the very attractive efficiency of the SOFC/GT hybrid systems, e.g., 65–70% for atmospheric cells and 74–76% for pressurized cells [62]. What is really doubtful is whether they will be competitive with

gas turbine-based systems for large-scale electric power generation considering their low power density, e.g., less than 5 kW/m² (cf. ~3,000 kW/m² for a large GT), and high cost, e.g., about \$1,500/kW to \$3,000/kW in 2010 dollars [63] (cf., \$350/kW to \$500/kW for a GT depending on its size).



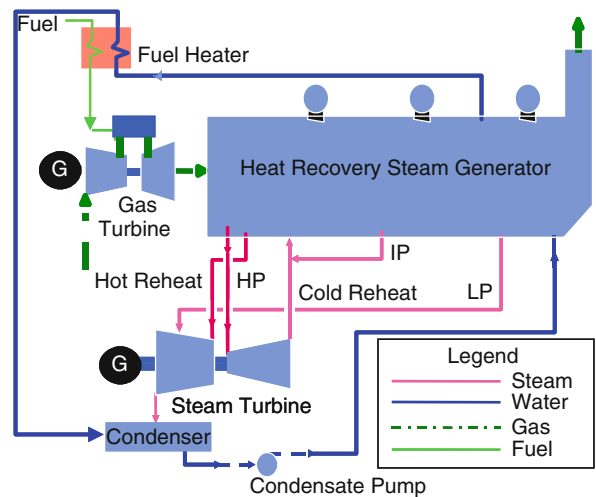
Natural Gas Power. Figure 16

“Real” combined cycle technology curves for different PR-RIT combinations. They are based on the gas turbine technology curves in Fig. 12. (Also shown is data from Table 4 (as published in Ref. [6]))

Today’s SOA in CC Power Plants

A schematic diagram of a CC system is shown in Fig. 17, which depicts the four major components: gas turbine, steam turbine, condenser, and the heat recovery steam generator (HRSG). Not shown is the cooling tower (CT), which cools the condenser cooling water. Usually, the heat rejection system (including condenser, CT and circulation pump), and feed pumps along with myriad smaller heat exchangers (such as the GT fuel gas heater in the upper left corner in Fig. 17) are collectively referred to as the “balance of plant” (BOP).

The equipment that combines the GT (i.e., Brayton cycle) and ST (i.e., Rankine cycle) is the HRSG. Hot GT exhaust gas ($>1,100^{\circ}\text{F}$ for modern F-Class units) transfers heat to condensate and generates superheated steam, typically at multiple pressure levels, which is used in the ST for power generation. Current SOA is a three-pressure system with reheat (i.e., heating of HP steam turbine exhaust, referred to as “cold reheat,” to the same temperature as the main or HP steam) with steam temperatures of $1,050^{\circ}\text{F}$ or higher. Highest steam pressure at the ST inlet (throttle) is around 1,800–2,500 psia, beyond which the pressures at the



Natural Gas Power. Figure 17

Three-pressure, reheat (3PRH) combined cycle diagram (Source: GE Energy [64])

generation point in the HRSG exceed the critical pressure and one would require a “once-through” steam generation system as opposed to the existing “drum type” system. From the fundamental thermodynamics discussed earlier, the theoretically maximum (net)

power output that can be obtained from the Rankine bottoming cycle of the CC plant is exactly equal to the exergy of the GT exhaust gas [55]. Today's most advanced plants with advanced F-, G-, and/or H-Class GTs and 3PRH steam systems, are capable of achieving ~72% of that theoretical maximum, which can only be achieved in a Carnot engine [56]. For rapid estimations, this translates into an ST power output that is about 50% of the combined power output of all GTs in the plant.

As the fundamental CC efficiency relationship, Eq. (23), and the graphical representation in Fig. 13 make it very clear, the cycle's heat rejection temperature is of prime importance. This temperature is dictated by the steam condenser's operation pressure, which is significantly below the atmospheric pressure. For the best performance, lowest possible steam condensation pressures, around 1.5 in. of Hg or even lower, are imperative. Achieving those low-pressure levels requires striking a delicate balance between heat rejection equipment size, cost, type, and site ambient conditions. The key heat rejection system equipment is the steam condenser. There are many types of condensers that are used to handle the heat rejection from the CC plant. Depending on the site ambient conditions, cooling medium availability, environmental regulations and customer-specific economic criteria, a suitable choice can be made. The major options are listed below.

1. Water-cooled surface condenser
 - Once-through or open-loop (utilizing a naturally available cooling water source such as lake, river, ocean)
 - Closed-loop with a natural (very rare for CC plants) or mechanical (forced or induced draft) cooling tower (CT)
2. Dry, air-cooled (also known as the "A Frame") condenser
3. Air-cooled wet surface (hybrid) or "plume abatement" condenser
4. Direct contact condenser with dry CT (Heller) system

In general, water-cooled systems, in particular open-loop systems near a relatively cold natural source of cooling water (e.g., in coastal locations in the northern hemisphere), offer the best performance, i.e., the lowest condenser pressure. How low in condenser

pressure one can go is essentially a question of plant economics. Recently, permitting requirements in conjunction with environmental considerations and water scarcity in many geographic locations render water-cooled heat rejection systems infeasible. Significant size and cost of "dry" or air-cooled systems preclude the optimal condenser pressures and optimistic performance projections displayed in Fig. 16.

The key to affordable and sustainable large-scale electric power generation is economies of scale that can be achieved by ever-larger prime movers (gas and steam turbines of 300+ MW) and plant blocks. Modern CC plants include one or more GTs with a single ST; either in a "multi-shaft" or "single-shaft" configuration (e.g., see Figs. 18 and Figure 19). The most common multi-shaft block is 2×1 (i.e., 2 GTs and 1 ST) with power ratings of 600–900 MW. A few larger plants in 3×1 or 4×1 configuration or in multiple blocks of 2×1 are also available (e.g., see Fig. 18). Single-shaft (i.e., 1×1) plants with power ratings of 300–500 MW per "block" can also be combined into very large generating stations via multiple blocks. For a representative selection of modern CC power plant configuration options and performances, the reader is referred to the white papers listed in [64–68].

Even better economies of scale are achieved via the "reference plant" approach, in which OEMs develop pre-engineered modules that can be shipped and installed anywhere in the world with minimal customization to accommodate different site and customer requirements. This ensures that natural gas burning GT-CC power plants are not only the most efficient means of generating electric power available today, but also the most cost-effective (e.g., at a turnkey capital cost of \$600–1,000/kW and cost of electricity of a few cents per kWh) and fastest to build (e.g., less than 24 months of construction time). On top of all that, GTs with modern "dry low NO_x" combustors are also the cleanest fossil fuel burning technology in terms of emissions. This aspect will be explored in detail in the following section.

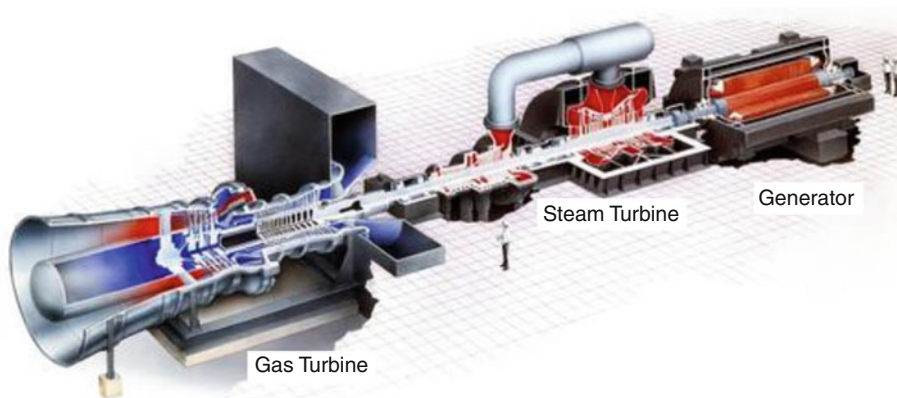
Cogeneration

Discussion thus far has focused on production of electricity from natural gas. When electricity is the sole product from the thermal cycle the heat rejected to



Natural Gas Power. Figure 18

Illustration of a $2 \times 207\text{FA}$ (two Frame-7 60-Hz GTs with single ST) NG CC power plant. From *left to right*, for each GT-HRSG “block” one can see the GT inlet air filter housing and duct (on top of the GT enclosure), HRSG with drums on the top, and the stack. Two STs are located to the right of the GT-HRSG blocks (cross-over pipe and down-flow LP section are clearly visible). Note the elevation of the ST structure to accommodate the steam condenser. Partially visible to the right is the multicell mechanical cooling tower. In the lower right corner are two distillate oil (GT backup fuel) tanks (Courtesy: GE Energy)



Natural Gas Power. Figure 19

A single-shaft combined cycle train (Courtesy: GE Energy)

ambient is termed “waste heat” because it is no longer economically useful for generation of additional power. In the larger context of a modern industrial society one finds that low temperature heat of the quality least

useful for power generation has many economically important uses. It is thus common in industrial and urban settings to integrate a power plant with another process such that heat energy for that process is

provided in whole or in part by the power cycle after the highest quality fuel energy has produced some useful work. Such installations are termed cogeneration systems or Combined Heat and Power (CHP) systems and can be very efficient from an environmental and fuel use perspective. Table 5 provides a sampling of cogeneration opportunities with a synergistic combination of heat and power needs.

In the ideal case, the heat energy available from the prime mover (engine or thermal cycle) will always match the needs of the customer process. In practice, this is challenging since many process heat demands are variable (such as building or district heating). The electrical loads can also be out of phase with process heat demands (e.g., an electric grid with high but unsteady wind or solar renewable power contributions). These variations in electrical load vs. process heat loads can be accommodated with appropriate system design and operating strategies. A variety of cogeneration plant options are available between the extremes of a non-cogeneration facility, e.g., fired heater or boiler, with no power generation at all (e.g., a home heating system capturing the heat content of the fuel but wasting its potential to do work [69]) and the aforementioned power generation combined

cycle with no process heat demand, which has accordingly been optimized for production of electricity.

Primary considerations in cogeneration system design are the size of the plant, the relative need for heat vs. electrical power, and the electricity and process heat demand variability. The following points summarize the range of system design solutions:

- Large industrial gas turbines ranging from ~20 to 300 MW electrical output are optimized for best combined cycle efficiency, and have exhaust energy of about 60% of heat input with exhaust temperatures between about 1,020°F and 1,160°F (550°C and 625°C). This class of machine is well suited to process heat demands that are large in an absolute sense but small or intermittent in relation to electrical power demand, (0% to ~30% of gas turbine heat input). This level of heat energy supply can readily be accommodated by modifying the power generation combined cycle system to include the capability to extract or divert steam from the bottoming cycle steam turbine to the process. This is a common solution for district heating which may be highly variable and seasonal.
- Aeroderivative gas turbines ranging from ~5 to 60 MW electrical output are optimized for best simple cycle efficiency, and have exhaust energy of about 55% of heat input with exhaust temperatures of about 850°F (455°C). These machines are large enough to support substantial process heat loads and are often configured in a system without a steam turbine to save cost and make a higher portion of gas turbine exhaust energy available to process. In many industrial cogeneration facilities, it is more important to match process heating demand than to meet a specific electric load. The gas turbine load is thus varied to follow process energy demand independent of electric load. Any excess electricity generated is sold to the grid and any power shortfall is purchased. If the design process heat demand is high in relation to electric load, or may occasionally need to be much higher than the unfired gas turbine exhaust energy, the system will need to include a supplemental burner in the HRSG, which can be used to boost process heat export capability. A burner may also be used in combination with part load operation of the gas

Natural Gas Power. Table 5 A sampling of natural gas Fired cogeneration opportunities [71]

Industrial	
Paper and board manufacture	Pharmaceuticals and fine chemicals
Ceramics	Brewing, distilling, and malting
Textile processing	Food processing
Oil refineries	Minerals processing
Timber processing and paper	Horticulture and glasshouses
Buildings	
Hotels	District heating and cooling
Airports	Hospitals
Supermarkets and large stores	College campuses and schools
Individual houses	Office buildings

turbine to maintain process heat export while reducing electricity production to match its load.

- A micro gas turbine ranging from ~ 0.03 to 0.3 MW electrical output may be selected if the heat load and/or electric load is small. Micro turbines typically employ uncooled regenerative cycles for best simple cycle efficiency, have exhaust energy of about 75% of heat input, and exhaust temperature of about 530°F (275°C). These units are too small to economically justify a bottoming cycle steam turbine so that they fall into the unfired or supplementary-fired heat recovery-type system configuration.
- In the size range between micro turbines and aeroderivatives, a variety of small industrial gas turbines and other prime movers such as a spark ignition reciprocating engine, or fuel cell, may also be considered. In the range from ~ 10 to 150 MW, direct-fired steam Rankine systems with steam

turbine extraction for process heat can also be an attractive configuration that is well suited to highly variable process heat needs. Power can be varied across a wide range with modulated fuel consumption and heat export can also be widely varied via controlled steam extraction. Efficiency at low process demand is inferior to gas turbine-based systems but fuel flexibility is generally better.

Table 6 shows some examples of gas turbine-based natural gas-fired cogeneration systems. These are presented to elucidate the behavior of such systems and draw conclusions regarding the role of cogeneration in promoting sustainable use of natural gas. Columns tabulate net electric power export, export heat energy to power ratio, quality of export heat expressed as a ratio of its availability to produce work (exergy), electric generation efficiency, cogeneration efficiency expressed on both an energy and exergy basis, and

Natural Gas Power. Table 6 Typical heat-to-power ratios and efficiencies for GT-based CHP systems

							Engine+Burner CO ₂ Emissions	
	Net power (MWe)	θ	β	η_{el} (%)	$\eta_{c,I}$ (%)	$\eta_{c,II}$ (%)	A	B
1. Unfired micro GT	0.1	1.74:1	0.29	25.4	69.6	38.3	284.7	516.9
2. Supplementary fired micro GT	0.1	10.0:1	0.44	8.6	93.0	46.6	212.9	418.3
3. Unfired aeroderivative GT	40	1.13:1	0.41	41.2	87.8	60.4	225.4	327.7
4. Supplementary fired Aeroderivative GT	40	2.63:1	0.44	26.3	93.0	56.6	212.9	349.6
5. Unfired industrial GT	100	1.50:1	0.45	35.6	89.0	59.5	222.5	332.8
6. Supplementary fired Industrial GT	100	2.84:1	0.44	24.8	93.0	55.7	212.9	355.4
7. Gas turbine combined cycle (unfired, industrial GT)	100	0.1:1	0.29	54.3	59.7	55.9	331.8	354.4
		0.5:1	0.34	49.1	73.6	57.4	269	344.7
		1.0:1	0.38	42.4	84.9	58.7	233.3	337.2
8. Gas turbine combined cycle (industrial GT with supplementary firing)	100	1.5:1	0.44	35.2	88.0	58.5	225.1	338.8
		2.0:1	0.44	31.6	93.0	59.4	212.9	333.2

θ is heat to power ratio in kWth/kWe

β is heat quality, which is the ratio of export exergy to export energy

η_{el} is electric generation efficiency

$\eta_{c,I}$ is first law cogeneration efficiency (energy-to-fuel input)

$\eta_{c,II}$ is second law cogeneration efficiency (exergy-to-fuel input)

Emissions are in g/kWh of energy (A) and exergy (B)

specific CO₂ emissions also expressed on both an energy and exergy basis. Export heat is provided as steam in all cases.

Cogeneration efficiency $\eta_{c,I}$ in Table 6 is the straight ratio of the sum of power (kW_e) and heat (kW_{th}) products to the fuel burned in the plant on an LHV basis. This is a commonly used (sometimes with fuel input on an HHV basis) measure, which is ultimately misleading because it does not differentiate between the “quality” of different products. In certain cases, especially with supplementary firing, it can reach improbably high values such as 90% or more.

A more rational measure is the cogeneration efficiency per the second law of thermodynamics, $\eta_{c,II}$, which uses the exergy of the thermal product rather than its energy (i.e., enthalpy) [70]. Exergy, as discussed earlier in the present article, is a direct measure of the work generating ability of a material stream via Kelvin–Planck statement of the second law of thermodynamics. Exergy is a property of a given fluid and, as such, for two given properties (e.g., pressure and temperature), it can be calculated using a suitable equation of state such as ASME steam tables. The relationship between $\eta_{c,II}$ and $\eta_{c,I}$ is given as

$$\eta_{c,II} = \eta_{c,I} \cdot \frac{1 + \beta \cdot \theta}{1 + \theta}$$

where θ is the heat-to-power ratio of the cogeneration plant, and β is the ratio of the exergy to energy (enthalpy). Since $\beta < 1$, $\eta_{c,II} < \eta_{c,I}$ and has typically a value, which is more in line with standard power plant efficiencies. In passing note that the PURPA (*Public Utility Regulatory Policies Act*, a US federal law enacted in 1978 intended to encourage more energy-efficient and environmentally friendly commercial energy production) efficiency, which is another cogeneration plant efficiency definition, is equivalent to $\eta_{c,II}$ with $\beta = 0.5$. Even $\eta_{c,II}$ is an inflated value because β is the efficiency of a hypothetical Carnot engine that converts the thermal energy to useful shaft work.

A realistic turbine that can utilize that thermal efficiency for electric power generation would have a much lower efficiency, which is characterized by the rational efficiency, ε , which is the ratio of actual turbine power output to that of the hypothetical Carnot engine. The most common thermal energy product in cogeneration is steam. Depending on the pressure and

temperature of steam, with today’s state of the art in ST technology, a reasonably good estimate for ε would be around 0.7. Thus, a more realistic value for $\eta_{c,II}$ would be given as

$$\eta_{c,II} = \eta_{c,I} \cdot \frac{1 + \varepsilon \cdot \beta \cdot \theta}{1 + \theta}$$

Case 1 is a micro turbine–based design with no supplemental fuel addition to exhaust heat recovery. Export energy is high in relation to power production since the engine is only 25% efficient but export heat quality is low since exhaust temperature is a modest 530°F (275°C). This low exhaust temperature also compromises the cogeneration energy efficiency in these examples because only a portion of the unfired exhaust energy is recoverable to steam.

Case 2 adds a supplemental burner fired to 1,600°F (871°C) to boost export energy substantially while also improving its quality. Fuel heat energy is now fully utilized to provide power and process heat (93% cogeneration efficiency, energy basis), but the performance is still modest on an exergy basis since the engine is inefficient. Note also that specific CO₂ emissions (exergy basis) improved because the incremental efficiency due to increased and improved heat export is better than the electrical efficiency of the gas turbine.

Cases 3 and *4* show the corresponding story for a 40 MW aeroderivative gas turbine while *cases 5* and *6* present a 100 MW industrial gas turbine scenario. These gas turbines both improve exergetic performance over the micro turbine cases but when supplemental firing is added the cogeneration efficiency and specific CO₂ emissions (exergy basis) both degrade. The more efficient aeroderivative has slightly higher system performance but lower process heat export capability vs. the industrial gas turbine.

Case 7 shows the 100 MW industrial gas turbine in a combined cycle configuration with increasing steam extraction to process. Electric generation efficiency falls off as more energy is diverted to process while cogeneration efficiencies improve. *Case 8* begins where *case 7* left off with power output maintained as supplemental firing is increased. By the time 1,600°F (871°C) supplemental firing

temperature is reached the export heat energy has approximately doubled. As supplementary firing increases the specific CO₂ emissions (exergy basis) stays approximately constant.

Note that the supplementary-fired scenarios presented here stop at 1,600°F (871°C) consistent with uncooled metal ductwork. Firing temperature can be pushed further by application of refractory or water-cooled walls until the excess oxygen is nearly expended; or increased even further if supplemental air is also added. High supplementary firing may be required for some applications but is less desirable from a resource conservation perspective than a design with little (or no) supplementary firing.

The following general conclusions can be drawn from this sampling:

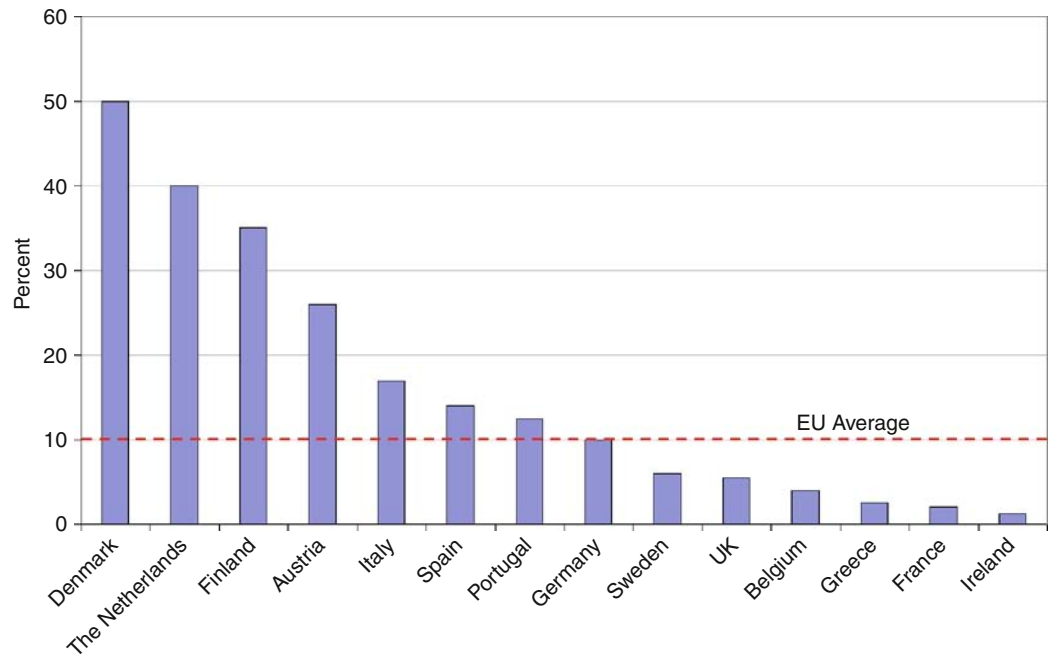
- 1. Cogeneration efficiency should be expressed on an exergy basis to promote the most sustainable use of fossil fuels. Direct use of fuel for low-quality end uses represents a large waste of the thermodynamic potential inherent in the fuel. Cogeneration

efficiency as commonly expressed in relation to export energy gives the false impression that supplementary firing improves the environmental (CO₂) and resource (natural gas) utilization performance.

- 2. Cogeneration system configuration will be a function of the quality and quantity of heat energy required, its variability, and the relative heat and power needs. A wide range of applications can be served by appropriate system design.
- 3. Gas turbine-based systems are most efficient when not supplementary fired while gas turbine combined cycle cogeneration systems maintain high cogeneration exergetic efficiency even with supplemental firing. This is because fuel energy can be recovered to power all the way down to the temperature level required by the process user.

Figure 20 shows cogeneration’s contribution to electricity production in European Union countries.

The higher penetration seen in four of these countries is attributed to a combination of favorable economics and supportive government energy policy [71].



Natural Gas Power. Figure 20

Cogeneration as a share of national power production (1999) (Source: European Association for the Promotion of Cogeneration [71])

Natural Gas Power. Table 7 Annual US thermal energy available supply and rejection, 2008 [72, 73]

	Coal	Petroleum	Natural gas			Nuclear	Total
			Combined cycle	Simple cycle	Cogeneration		
Electricity generation (Billion kWh)	1,986	46.2	562	122	200	806	3,721
Heat rate (Btu/kWh, HHV)	9,884	9,884	7,598	11,526	11,156	10,488	
Heat input (trillion Btu)	19,628	457	4,268	1,402	2,227	8,456	36,062
Approximate heat rejection to ambient or process (trillion Btu)	10,889	254	1,408	847	1,159	5,451	19,730
Heat rejection beta (β)	0.20	0.20	0.20	0.44	0.44	0.20	

In the USA, approximately 4% of electricity production currently comes from cogeneration facilities of which 71% are fueled by natural gas (e.g., see Table 8.2c in [72]).

How far could cogeneration penetrate in the USA and what benefits would accrue? Just as it is possible to look at the utilization rate of renewable energy sources such as wind, hydro, and solar in relation to the size and distribution of these resources, the thermal power plant heat available for low-temperature end users can be quantified to estimate the potential for increased cogeneration and hence its potential to offset other fuel use and emissions.

The annual US thermal power generation totaled 3.7 trillion kWh in 2008 [72]. Factoring in the mix of natural gas-fired systems, nuclear, and direct-fired (mostly coal) plants and their respective thermal efficiencies allow estimation of heat resources available for cogeneration. Table 7 shows an estimated breakdown of heat rejection from thermal power plants in the USA.

While cogeneration with direct-fired or nuclear-fueled Rankine systems is certainly possible, the sheer size of these plants makes it difficult to economically match them to end users that themselves are large enough to benefit from the full heat energy available. For this reason, cogeneration systems tend to be relatively small in terms of electrical output, which in turn aligns well with the NG-fired gas turbine and gas turbine combined cycle. Natural gas accounts for 883 billion kWh or 23.7% of total electricity generation [72, 73]; 199.7 billion kWh or 22.6% of which comes from cogeneration facilities (e.g., see Table 2.1 in [73]). Taking the data in Table 6 for cases 5 and 6 as a typical

system configuration, it is possible to estimate the current fuel savings from natural gas-fired cogeneration as follows.

Natural Gas-Fired Cogeneration (GT+Supplemental Firing)

1	Electricity generation (billion kWh, 2008)	200
2	Combined cycle heat rate (Btu/kWh, HHV)	8,305
3	GT heat input (trillion Btu, 2008)	1,658
4	System heat rate (Btu/kWh, HHV)	11,156
5	Burner supplemental heat input (trillion Btu, 2008)	569
6	System heat input (trillion Btu, 2008)	2,228
7	Heat export to process (trillion Btu, 2008)	1,159
8	Equivalent heat input to meet process (trillion Btu, 2008)	1,302
	Cogeneration natural gas savings (trillion Btu, 2008) (Line 8 minus Line 5)	733

The fuel savings from application of cogeneration represents ~24.8% reduction in fuel burn vs. making the power with a gas turbine combined cycle and supplying the heat independently with a fired boiler. CO₂ emissions reductions follow the fuel saving directly if the fuel saving is booked as natural gas. If booked at the current fuel mix for US electricity production, the CO₂ savings are higher by about 33% or higher by 75% if booked as reduced coal fired generation. Looking at the heat rejection from natural gas-fired power generation in the USA in Table 7, in comparison to current heat

export from natural gas-fueled cogeneration, it appears that the amount of cogeneration could perhaps be doubled without a major shift in fuel use patterns. Reaching the electricity contribution levels seen in some small European countries would likely require more distributed generation with replacement of large coal-fired power plants with more small cogeneration systems (likely gas fired) proximate to potential end users.

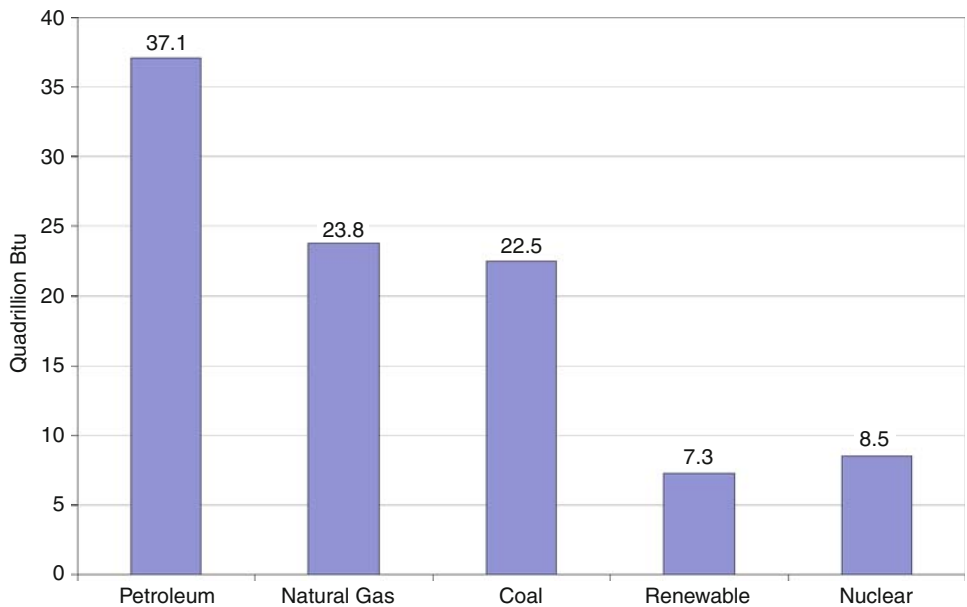
Turning now to the potential customers for heat energy, Fig. 21 shows the primary energy production by source in the USA for 2008 [72]. Figure 22 shows the incremental retail electric load distributed to each sector and gives a breakdown between residential and commercial sectors. The natural gas contribution to the residential and commercial sectors is 8.2 trillion Btu [72]. Assuming that its primary use is for space heating, it can be seen that it far exceeds the available waste heat from gas-fired power plants tabulated in Table 7. It is difficult to make the case that small (micro) cogeneration systems could be aligned to the

heat and power needs at the individual house level, though work on fuel cell systems for this purpose is active. The commercial sector and high-density residential (e.g., apartment complex) does offer real potential for expansion. In either case, the performance of the cogeneration system needs to be high enough to improve on the fuel use and environmental performance of an efficient utility scale power plant in combination with direct use of natural gas for heating.

Combustion and Emissions

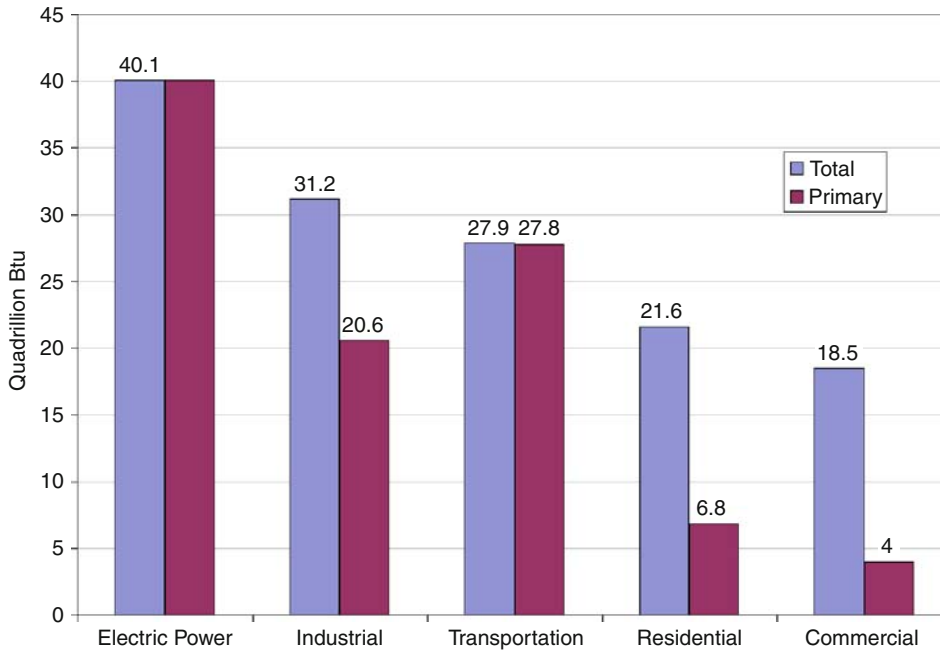
There are different types of GT combustors that have been used at one time or another by different OEMs in their products:

- 1. Silo (obsolete now, once used by Siemens and former ABB)
- 2. Annular (aero-derivatives such as GE LM6000)
- 3. Can or tubular (heavy-duty industrial GTs by Siemens, GE)



Natural Gas Power. Figure 21

Primary energy production by source, 2008. Total primary consumption is 99.3 quadrillion Btu. Renewable energy sources are hydroelectric power, geothermal, solar/photovoltaic, wind, and biomass. "Petroleum" does not include the fuel ethanol portion of motor gasoline – fuel ethanol is included in "Renewable Energy." "Natural Gas" excludes supplemental gaseous fuels. "Coal" includes less than 0.1 quadrillion Btu of coal coke net imports (Source: Energy Information Administration (EIA) [72])



Natural Gas Power. Figure 22

Primary and total consumption by sector, 2008. See Fig. 21 for a breakdown of sources (Source: Energy Information Administration (EIA) [72])

In the following discussion a parameter will appear frequently, i.e., the *equivalence ratio*, ϕ , which is the ratio of the actual fuel-air ratio, f , to the *stoichiometric* fuel air ratio. The latter is the theoretical value requisite for complete combustion. Thus,

$$\phi = f_{\text{act}} / f_{\text{st}}$$

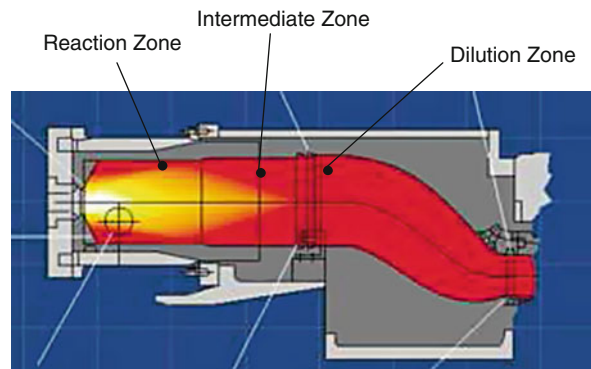
When the equivalence ratio is equal to 1.0, the flame temperature is the highest and the chemical reactions are the fastest. Thus,

1. If ϕ is less than 1.0, the mixture that reacts is referred to as “lean” or “fuel lean.”
2. If ϕ is greater than 1.0, the mixture that reacts is referred to as “rich” or “fuel rich.”

In either lean or rich combustion, flame temperature is lower and chemical reaction is slower than their stoichiometric counterparts.

The can-type combustor can be divided into three zones (see Fig. 23):

Primary Zone (PZ): This is the combustor *reaction* or flame zone where fuel is injected and mixes with air (about 15–20% of the compressor discharge air is



Natural Gas Power. Figure 23

Single-Nozzle Diffusion (SND) Combustor

introduced). In order to cap the combustor exit temperatures at a level, which downstream turbine HGP components can tolerate, the requisite value of ϕ is typically 0.4–0.5, which is too lean for stable and efficient burning. This is why only a portion of the combustion air is introduced into the primary zone. High-temperature rapid combustion takes place in the PZ. The design of this part of the combustor must

facilitate highly turbulent and recirculating flow pattern to anchor the flame at the exit of the fuel injector nozzle. This is critical because of the high-speed of the air stream, even at a fraction of its full value, which is an order of magnitude higher than the turbulent flame speed at which the flame propagates into the fuel-air mixture.

Secondary Zone (SZ): This is the *intermediate* zone where a portion of the remaining combustion air (~30% of the total) is introduced in the mid-section of the can through holes or cooling slots in the combustor liner. The hot combustion gas from the PZ gets thermally soaked and the combustion is completed. The air that flows between the liner and flow sleeve and enters the can forms a film to cool the liner wall.

Dilution Zone: This is where the remaining air enters the can through the metering holes and cools the gases to the desired temperature. A design to promote turbulence and thorough mixing of hot and cold streams is essential to ensure a homogeneous temperature distribution and prevent hot streaks.

A successful combustor design should satisfy the following criteria:

1. Efficient and stable operation, which comprises the following characteristics
2. Complete combustion at or near stoichiometric flame temperature
3. Flame stability and reliability at high turndown
 - (a) Ignition
 - (b) Crossfire (propagation of flame from fired chambers with spark plugs to the unfired chambers via tubes interconnecting the annular combustion chambers or cans)
 - (c) Lean blowout
4. Operational stability and reliability at high turndown
 - (a) Dynamics (noise), i.e., large-amplitude pressure oscillations in combustion chamber, driven by heat release oscillations; also known as combustion “humming.” These oscillations are destructive to engine hardware.
 - (b) Pattern factor, which is a measure of the temperature distribution at the combustor exit defined as $PF = \frac{T_{\text{peak}} - T_{\text{mean}}}{T_{\text{mean}} - T_{\text{inlet}}}$
 - (c) Pressure drop
5. Low emissions in compliance with regulatory laws

Achieving these goals simultaneously is very difficult due to the counter-acting impacts of design parameters on each criterion. The design history of the GE heavy-duty industrial gas turbine combustors can be found in Ref. [74]. The design evolution can be summarized in three major designs:

1. Single-nozzle diffusion (SND) combustor
2. Multi-nozzle, quiet combustor (MNQC)
3. DLN or Dry-Low-NO_x combustor

The design evolution was primarily driven by the desire to satisfy the increasingly stringent emission control requirements with improved efficiency via higher combustion and turbine inlet temperatures (TIT). In order to clarify this, one must be familiar with the chemical mechanisms of NO_x production, which is one of the most critical pollutants present in GT exhaust gas. Principal pollutants emitted by a GT are as follows [75]:

1. Oxides of nitrogen (NO_x), which are toxic, contribute to chemical smog and lead to depletion of ozone in stratosphere
2. Oxides of sulfur (SO_x), which are toxic and corrosive
3. Carbon dioxide (CO₂), which contributes to greenhouse effect and global warming
4. Carbon monoxide (CO), which is toxic and asphyxiating
5. Unburned hydrocarbons (UHC), including volatile organic compounds (VOC), which contribute to urban smog
6. Particulate matter (PM), which is suspected to cause respiratory diseases

Detailed information on GT emissions and control can be found in Ref. [76]. Some pollutants, i.e., UHC, PM, and SO_x, are generally of no concern for GTs burning gaseous fuel with the exception of coal-gasification SG, which leads to conversion of fuel-contained sulfur to SO_x. (Note, however, that in some cases contamination of fuel in the pipeline, e.g., rust particles, can be a problem for HGP components.) In any event, currently there is no available GT combustion technology to prevent or control SO_x emissions. Similarly, CO₂ is a direct product of the complete combustion process and cannot be controlled or reduced within the GT control volume (except, of

course, increasing the thermal efficiency of the GT, which will reduce the amount of fuel consumed and CO₂ generated for a given power output). Precombustion or postcombustion capture and storage of carbon dioxide, known as carbon capture and storage (CCS), is currently a hot topic of research and development.

A key pollutant emitted by the GT is NO_x. There are two major sources of NO_x:

1. Atmospheric N₂ in post-flame gases, which leads to formation of NO and NO₂ via three principle mechanisms:
 - (a) Thermal NO (extended Zel'dovich mechanism)
 - (b) Nitrous oxide mechanism
 - (c) Prompt NO mechanism
2. Organically bounded nitrogen in fuel or fuel-bound nitrogen (FBN)

The dominant NO_x formation mechanism is the thermal or Zel'dovich mechanism. Although not a significant source of NO_x for NG, FBN in low-quality liquid fuels and LCV gases with hot-gas clean-up (via ammonia) can be high enough to merit attention. NO_x (and CO) formation is a function of three primary parameters:

1. Residence time in the combustion zone
2. Chemical reaction rate
3. Mixing rate

These three reaction parameters can be related to the turbine operating conditions and combustor size [75]. In particular,

$$\text{Residence time} \propto \frac{p \cdot V_c}{\dot{m} \cdot T_{pz}} \quad (27)$$

$$\text{Reaction time} \propto p^n \cdot \exp(kT_{st}) \quad (28)$$

where

- T_{st} is stoichiometric flame temperature (or, equivalently, f or ϕ or firing temperature)
- T_{pz} is the average gas temperature in the reaction or primary zone
- V_c is combustion volume
- p is combustion pressure (raised to a power of n)
- \dot{m} is the combustor mass flow rate

Therefore, NO_x production rate can be expressed in an empirical formula that can be written as

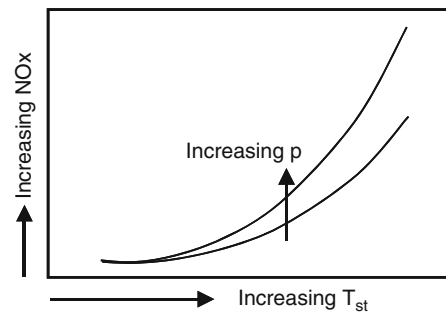
$$\text{NO}_x \propto \frac{V_c}{\dot{m} \cdot T_{pz}} \cdot p^{1+n} \cdot \exp(k \cdot T_{st}) \quad (29)$$

The relationship between NO_x production and key parameters per Eq. (29) is shown qualitatively in Fig. 24. In general, NO_x values are reported in parts per million (ppm) on a dry basis (usually expressed as ppmvd) and corrected to 15% O₂ in combustion air. It is extremely important to make sure that the basis of the reported number is clearly defined; otherwise comparisons of data from different sources can be misleading. The expression for correction is as follows:

$$[\text{NO}_x]_{15\% \text{O}_2} = \frac{[\text{NO}_x] \cdot 5.9}{(20.9 - \text{O}_{2,\text{meas}})} \quad (30)$$

The trends in Fig. 24 show that the impact of pressure on NO_x formation is stronger at high flame temperatures near the stoichiometric value (i.e., $\phi \sim 1.0$) and negligible at leaner mixtures with lower flame temperatures. Consequently, effective control of the NO_x production in the combustion process is a matter of reducing the flame temperature. There are essentially three ways to accomplish this:

1. Design a primary (reaction) zone with leaner fuel-air mixture
2. Injection of a diluent, i.e., water or steam, into the primary zone of the combustor to act as a “heat sink” and reduce the flame temperature



Natural Gas Power. Figure 24

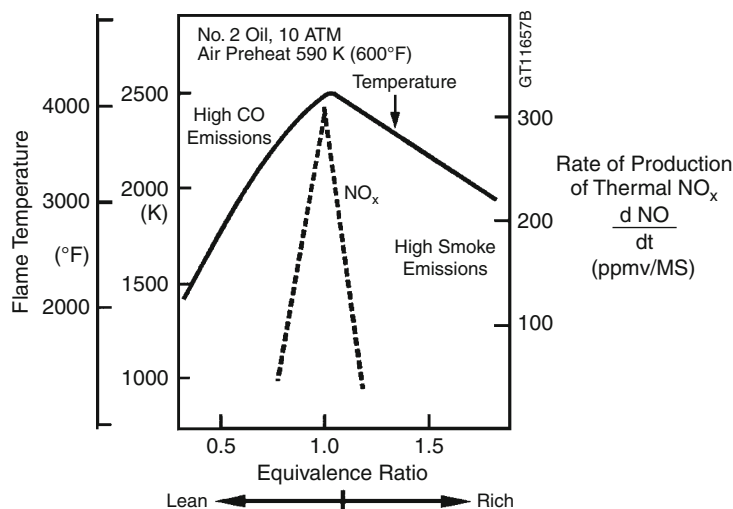
Effect of pressure and stoichiometric flame temperature on NO_x

3. Staged combustion with two distinct reaction zones, which lead to the modern “lean-premixed” designs, which is referred to as *Dry-Low-NO_x* (or *DLN*).

The dramatic impact of reducing ϕ , i.e., leaning out the PZ, on flame temperature and NO_x production can be seen in Fig. 25. Leaner fuel-air mixtures lead to lower-value flame and gas temperatures and favorably impact the NO_x production rate via shorter residence and reaction times.

Unfortunately, things get complicated rapidly. For starters, the full operating range of an industrial GT combustor requires significant turn-down ratios, e.g., 40-to-1 in fuel flow, 30-to-1 in air flow, and 5-to-1 in fuel-air ratio. Obviously, the combustor design must ensure stable and efficient combustion across the entire operating range. The early single-nozzle (diffusion flame) combustors, with direct injection of fuel and air into the PZ, were designed for combustion at or near stoichiometric conditions. For a combustor designed for optimal full load operation at $\phi \sim 1$, combustion at lower loads is bound to be very lean. Thus, further leaning out of the PZ for NO_x reduction is a very limited knob. In fact, initial efforts to reduce NO_x via leaning out reaction zone of an SND combustor did not achieve more than 20% reduction [77].

When it became clear that meeting increasingly stringent pollution-control regulations via leaner reaction zones in SND combustors, other methods to reduce flame temperatures were investigated. Injection of water or steam into the PZ to create a “heat sink” for temperature reduction proved successful. Starting in 1979, when regulations required that NO_x emissions be limited to 75 ppmvd, SND combustors with water or steam injection have been successfully used to meet or exceed these requirements successfully [74]. However, SND combustors that use water and steam injection are limited in their ability to reduce NO_x levels below 42 ppmvd on gas fuel (65 ppmvd on oil fuel). The main reason is the practical limit to the amount of diluent injection into the reaction zone before the increase in dynamic pressure oscillations (i.e., noise) start imposing unacceptable penalties in terms of reduced hardware life and increased maintenance frequencies. High levels of water injection excited discrete dynamic pressure tones within the combustor. Frequencies within the range of the hardware’s natural frequency resulted in combustor damage. In terms of dynamic pressure response, steam is a better diluent than water, although a larger mass flow rate is needed. Significant reduction in NO_x production is achievable via steam injection, e.g., ~60% at a steam-to-fuel ratio of unity [77, 78].



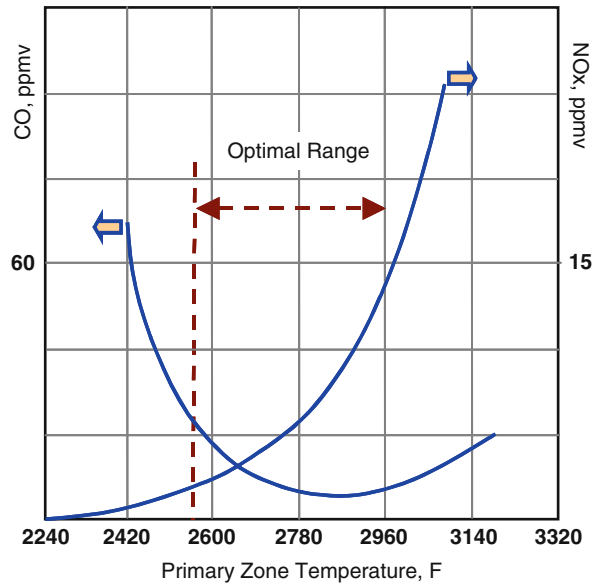
Natural Gas Power. Figure 25

Thermal NO_x production rate (Source: GE Energy [74])

The other significant drawback on water or steam injection for NO_x abatement is the overall simple or combined cycle performance penalty. Even though the turbine output increases via increased mass flow, the associated increase in fuel consumption is detrimental to the heat rate. Furthermore, since the quality of injection water must be high, i.e., of boiler quality, to prevent deposits and corrosion in HGP components, additional *water treatment system* (WTS) cost for simple cycle applications must be considered. In combined cycle applications, where steam from the HRSG or ST (typically *cold reheat* steam) is extracted for injection, increased make-up flow rates and ensuing WTS size increase and water consumption are additional concerns. Even disregarding the combustion noise and life issues, it is easy to imagine that ever-increasing turbine-firing temperatures for improved cycle efficiency would drive the diluent injection flows to such levels where they would literally put the flame out. This is notwithstanding the counterproductive effect of reduced cycle efficiency via higher fuel consumption, which defeats the purpose of increasing the firing temperature in the first place.

Another limit associated with lowering the equivalence ratio and flame temperature for NO_x abatement is the increase in CO emissions. The conflicting emission benefits and harms of lowering or increasing the flame temperature can be seen in Fig. 26. Carbon monoxide is a measure of combustion inefficiency, which becomes more pronounced at leaner combustions. This is due to shorter residence time at leaner mixtures, which does not lend itself to complete CO burnout. Thus, the allowable temperature range for emissions control is limited to a relatively narrow band of $\sim 400^\circ\text{F}$.

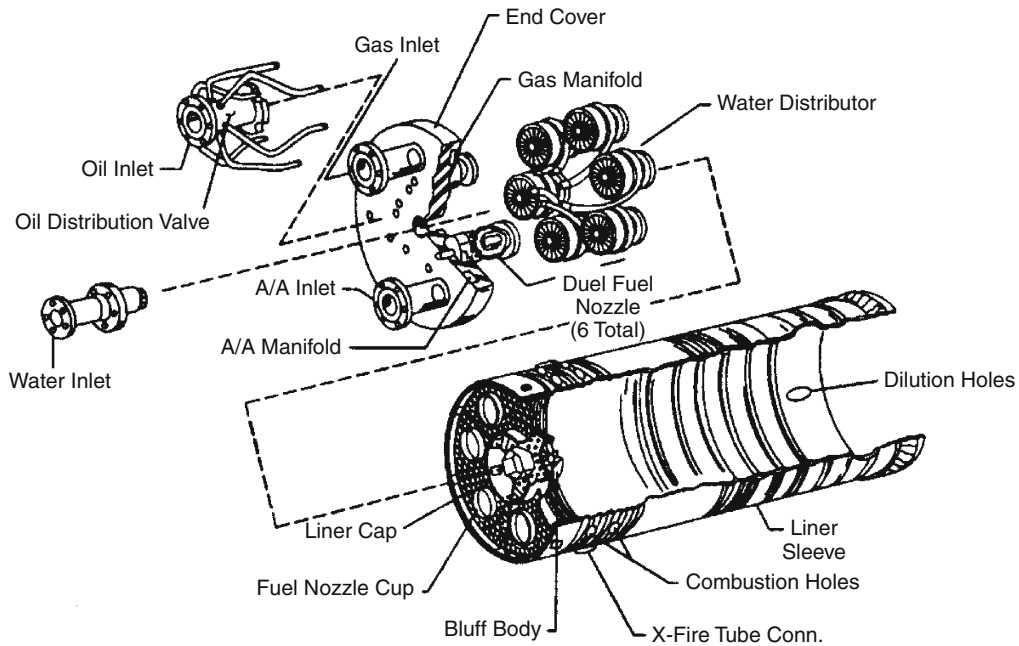
When SND combustors hit a practical limit for NO_x emissions reduction, even with diluent injection, designers responded by developing the “*multi-nozzle quiet combustion*” or MNQC system. For a detailed history of the combustor development, please consult Ref. [74]. Utilizing a multi-nozzle configuration, i.e., six nozzles located peripherally at the combustor head cap, this system eschewed the low-speed ignition and blowout problems associated with excessive leaning out of the PZ in SND systems at low loads (see Fig. 27). Furthermore, the MNQC system also helped with the dynamic pressure oscillation (i.e., noise) problem by



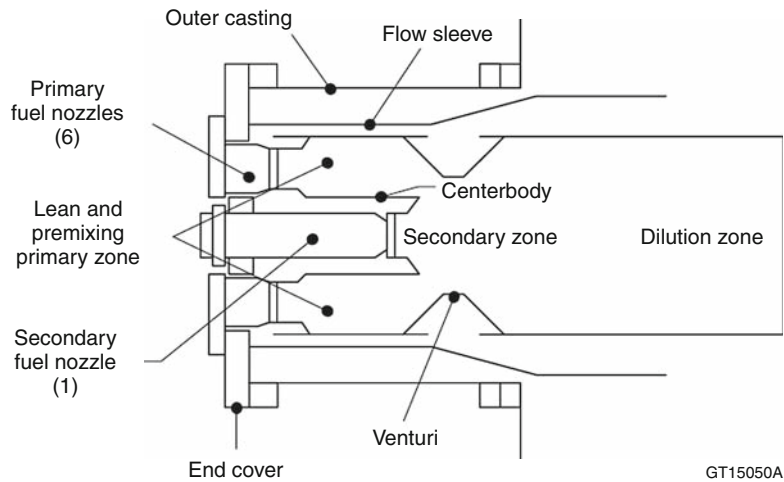
Natural Gas Power. Figure 26
Influence of PZ temperature on CO and NO_x [75]

maintaining a flat dynamic response at increased steam injection levels (that is the reason for the moniker “quiet”), thus enabling the design of combustion systems with higher firing temperatures and compliant NO_x emissions without the problems of increased hardware wear and reduced maintenance intervals. Since 1987, MNQC systems enabled heavy-duty GTs to meet stringent NO_x emission requirements, i.e., 25 ppmvd with gas fuels and 42 ppmvd with liquid fuels. Increase in CO and UHC emissions and combustion stability issues at high injection rates, e.g., steam-to-fuel >2 , precluded further improvement with this technology [80].

Reduction in NO_x emissions to even lower levels, the need for balancing the low emissions with ever-increasing firing temperatures and the need to avoid the cycle thermal efficiency penalty associated with steam and water injection lead to the development of DLN combustors. DLN combustion system is the natural next-step in combustor development, i.e., a “staged” combustion system with lean-premixed fuel-air mixture. The primary and secondary combustion zones in the staged combustor are operated at the lowest possible ϕ at high fuel flow conditions (see Fig. 28). During start-up, acceleration, and



Natural Gas Power. Figure 27
MNQC Combustor (Source: GE Energy [79])



GT15050A

Natural Gas Power. Figure 28
Dry-Low-NO_x (DLN) combustor (Source: GE Energy [80])

low-load conditions, fuel is introduced only in the PZ with good ignition and blowout characteristics. At increasing load levels, fuel is introduced into the SZ at a fuel split to balance the CO and NO_x production at an

optimal level, i.e., near the middle of the allowable range shown in Fig. 25. The first successful DLN system, DLN-I, that went into production incorporated a multi-nozzle system similar to that in MNQC for the

PZ and a conventional, large secondary fuel nozzle for the SZ. At normal (50–100% load) operation, fuel and air are “premixed” in the PZ but ignition takes place in the SZ with the flame attached to the secondary nozzle.

DLN technology successfully balanced the requirements for low combustion noise (i.e., dynamic stability) at part-load operation and optimal equivalence ratio for minimum NO_x at acceptable CO levels. DLN achieves this balance without diluent injection (and associated efficiency penalty and cost) while satisfying the requirement for low NO_x emissions, i.e., 25 ppm or lower on gas fuels. A history of the development of DLN-I and DLN-II (for F-Class turbines with high firing temperatures of 2,400+°F) systems can be found in Ref. [74] along with a detailed description of DLN operation modes.

Basic Economics

Natural gas-fired gas turbines and GTCC power plants are the most economic choices for fossil fuel-based electric power generation. This is clearly represented by the data in Table 8. The data in the table reflects 2004 prices. The numbers and the relative position of NG-fired plants should be expected to show a certain variation as a function of changing economic climate (i.e., inflation, demand for generating capacity,

absolute and relative fuel prices, labor shortage, raw material prices, etc.) of which the recent worldwide depression that started in the last quarter of 2008 is a very good example. Nevertheless, a dramatic shift in the economic advantages of the CC plants vis-à-vis other nonrenewable large central stations or renewable options (i.e., wind, solar, etc.), which cost nearly ten times per kW capacity and available only in small sizes amenable to distributed power generation, should not be expected near term. A very brief introduction to the key aspects of the CC plant economics, i.e., plant capital (turnkey) cost or price and the cost of generating 1 kWh of electric power, is therefore quite useful.

Turnkey combined cycle plant price levels from 2009 Gas Turbine World (GTW) Handbook are shown in Fig. 29. Note that listed GTW prices are in 2009 US dollars for a basic natural gas-fired combined cycle with gas turbine generator, unfired multi-pressure heat recovery boiler with no bypass stack, condensing multi-pressure steam turbine generator, step-up transformer, water-cooled heat rejection, standard controls, starting system, and plant auxiliaries. Obviously, depending on the scope of equipment, site-specific requirements, geographic location, and competitive market conditions these prices will vary considerably. Also excluded from the prices in GTW are the indirect costs and other commercial

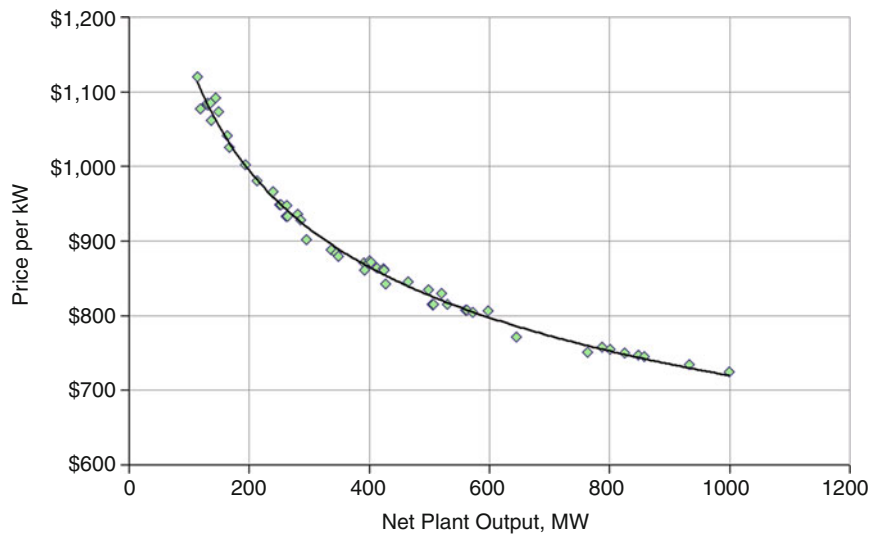
Natural Gas Power. Table 8 Cost and performance of nonrenewable electric power generation plants

	Capital cost per kW	Economic life	Lead time	Load factor	Efficiency		Fuel cost	Fixed	Variable
		years		%	%	% (CHP)	\$/MMBtu	\$/kW-year	\$/MWh
NG-fired GT	\$350–500	30	1.5	5–50	38			9–12	4
NG-fired GTCC	\$550–900	30	2	60–85	55		4–5	10–15	2
NG-fired GTCC with CHP	\$600–700	30	2	50–85	40–55	79–84		10–15	2
US SCPC	\$1,300–1,800	30	4	60–85	35–42			20–25	3
US IGCC	\$1,700–2,100	30	4	60–85	39–42		0.75–1.50	30–35	2
Nuclear	\$2,000–3,000	40	6	85–92	33		0.5	60–70	1.5

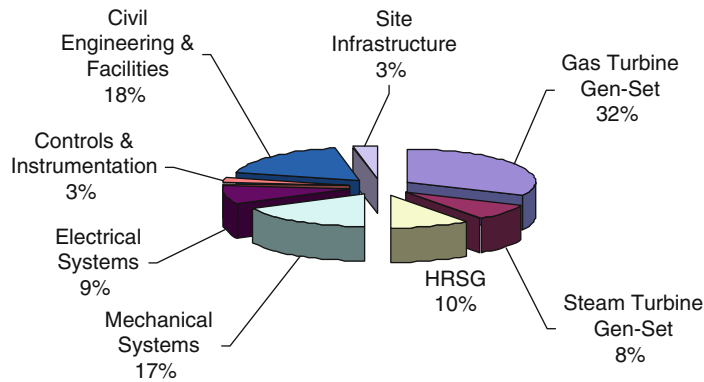
Source: Cambridge Energy Research Associates, MA, USA

Load factor specifies the percent of (equivalent) hours in a year (8,760 h) during which the plant runs at its nominal base load. The data reflects the 2004 prices

CHP Combined Heat and Power (also known as Cogeneration), SCPC Supercritical Pulverized Coal, IGCC Integrated (Coal) Gasification Combined Cycle



Natural Gas Power. Figure 29
Turnkey CC plant prices from Gas Turbine World 2009 Handbook. An adder of about 60% is applied to the listed price, which is budgetary turnkey equipment package price, to account for items not covered (see Fig. 30 and Table 9 below). Provided for reference only



Natural Gas Power. Figure 30
Direct cost breakdown for a typical 400 MW CCPP (Source: Reference [83], pp 16)

outlays. For a more realistic price level, the data in Fig. 29 reflects an adder of about 60% to the listed price, which is based on the assumption that the prices in GTW reflect only 75% of the direct costs, which are 85% of the total installed cost (see Fig. 30 and Table 9 below).

The data plotted in Fig. 29 can be summarized by the power law

$$k = 2,890 \cdot P^{-0.2012} \tag{31}$$

where k is the specific price (\$/kW) and P is the net plant output in MW (with $R^2=0.994$). In general, an exponent of 0.2–0.3 is adequate to estimate the capital cost of a CC plant from a known data point (i.e., price and net output). There are several methods and tools to calculate CC power plant cost (from a plant owner’s perspective), which are used within OEMs and A&E firms and not available for public consumption. Well-established cost-estimating techniques, especially those that are adopted by the CPI, are readily available in

Natural Gas Power. Table 9 Typical CC capital cost breakdown between direct and indirect costs. From Energy Issues (World Bank), No. 20 June 1999, which is based on a 350/700 MW CC Plant with a Siemens-Westinghouse V94.3A (now SGT5-4000 F) GT

Integrated services (Indirect costs)	Contribution to total CC cost (%)
Project management and subcontracting	4
Plant and project engineering// software	2
Plant construction, commissioning, and training	8
Transport and insurance	1
Direct costs	Contribution to total CC cost (%)
Civil works	15
Gas and steam turbine generators	32
Boiler island	11
BOP	16
Electrical systems	7
Instruments and control	4

textbooks and scholarly articles. For a discussion of those and the underlying details, the reader is referred to Reference [81]. Another valuable source is Reference [82].

Calculation of the final cost of the CC plant requires consideration of many elements. In fact, the total cost of GT and ST generators and the HRSG, which are the major equipments in the CC power plant, comprises about 50% of the total plant cost. A typical distribution is shown in Fig. 30 from Ref. [83]. In addition to these “direct” costs, there are “indirect costs” that can contribute anywhere between 15% and 30% to the final cost of the CC power plant. As an example, consider the representative breakdown in Table 9.

The presented data should suffice to give an idea about the scope of the CC plant capital cost and its importance from a design engineer’s perspective. All design performance calculations should be evaluated

with an eye to the final plant cost. That final plant cost is the money, i.e., the plant “price,” that is going to come out of the plant owner’s pockets or his credit sources and thus directly impact the generation cost of 1 kWh of electricity.

Probably the most critical customer criterion for choosing an electric power generation product is the life cycle cost of ownership. The metric that is used to quantify this cost is the unit (per kWh) cost of generating electricity. Cost of electricity (COE), in its most commonly used form, is a simple formula that combines a power generation system’s cost (capital and operating) and thermal performance into a single figure of merit that enables comparison of different systems. It is given as:

$$\text{COE} = \frac{\beta \cdot k}{H} + \frac{L \cdot f}{\eta} + \text{OM} \quad (32)$$

where β = Capital charge factor k = Total plant specific cost in \$/kW H = Annual operating hours f = First-year fuel cost in \$/MMBtu(HHV) L = Levelization factor η = Net efficiency of the CC plant OM = Operating & maintenance costs, \$/kWh.

The first term on the right-hand side of Eq. (32) represents the carrying charges, the second term is the cost of fuel, and the last term is the cost of maintaining and operating the plant. Fuel cost and operation and maintenance (O&M) costs are the variable components of the COE whereas carrying charge is the fixed component. COE is typically calculated over the economic lifetime of the plant (typically 30 years).

Levelization factor, L , converts the fuel cost that usually escalates over the plant’s economic life due to market conditions and inflation into average annual (constant) values. The formula for L can be found in any financial reference and is a function of discount, escalation, and inflation rates.

The fixed charge rate, β , represents the plant’s financial and tax situation and is a function of the economic life of the power plant, debt–equity ratio used in financing the plant, debt and equity rates, corporate tax rate and tax credits, book depreciation, property tax, and insurance. Various methods are available for calculating β , which can be found in plant economic studies. Typical values of β are 16–18% for *Independent Power Producers* (IPPs) and 12–13% for government-owned utilities, depending on the interest and taxation rates.

If the new and clean performance conditions are used within the calculations, the cost of generation as given by the COE formula should be treated as a figure of merit rather than an accurate assessment of the “real” generating cost. Important drivers – such as seasonal variations in ambient conditions, part load operation, recoverable and unrecoverable component deterioration, uncertainty in fuel prices, start-up and shutdown, and unforeseen events – are not considered. Nevertheless, levelized COE is an important metric that provides insight to performance-cost trade-offs for OEMs (system design) as well as customers (system selection).

The exact distribution of the cost of generating 1 kWh of electric energy varies depending on the financial assumptions. Typically, about two-thirds of the COE for an NG-fired, medium-loaded (~6,000 h/year) modern 3PRH CC plant with an F-Class gas turbine is the fuel cost. About one fourth of the COE is the capital cost. The distribution is markedly different (actually, nearly exactly the opposite of NGCC) for the IGCC plant with very high capital expenditure (\$2,000/kW or even higher) utilizing a cheaper fuel, coal (\$1–2).

Future Directions

At the time of writing of this article, natural gas accounts for about 22% of US electricity generation. The share of natural gas in electricity generation in OECD member countries in Europe is about the same. The share of natural gas is expected to increase in the future to the tune of 25% of total world electricity generation by 2030 [2]. However, recent developments suggest that this projection might in fact be a low estimate. One reason is the new drilling technologies that made extraction of natural gas from shale rocks feasible. The technique is called *hydraulic fracturing* and involves injection of high-pressure water into rock formations to destroy them and enable the trapped gas to be released to the surface for capture. So much so that this development resulted in a 35% jump in US natural gas reserves from about 1,500 trillion ft³ in 2006 to more than 2,000 trillion ft³ in 2008 [84]. Note that this number is the total estimated reserves including speculative, possible, and probable sources, and of which only 237 trillion ft³ is considered to be proven by the Energy Department. According to

the same source, shale gas constitutes nearly a third of the total reserves.

Another potential source of natural gas is *gas hydrates* (also known as *gas clathrates*), which are “crystalline solids consisting of gas molecules, usually methane, each surrounded by a cage of water molecules.” It looks very much like water ice. Methane hydrate is stable in ocean floor sediments at water depths greater than 300 m, and where it occurs, it is known to cement loose sediments in a surface layer several hundred meters thick. The worldwide amount of carbon bound in gas hydrates is conservatively estimated to total *twice the amount of carbon to be found in all known fossil fuels on Earth*. (Source: US Geological Survey Fact Sheet.) Unfortunately, as of today, commercially feasible production of natural gas from gas hydrates is not a reality.

The dramatic increase in natural gas reserves and actual production via new technology that can extract it from shale rocks, taken at its face value, is a strong driver for its increasing share in electricity generation; at least in the USA. The advantage of natural gas vis-à-vis coal, which roughly accounts for half of US electricity generation, in terms of reduced carbon emissions is thus fortified from lower prices and increased availability. But the situation may not be as simple as it seems. The price of natural gas hit its 7-year low of about \$3/1,000 ft³ in the summer of 2009. This huge drop from a peak of \$13 in the summer of 2008 was a direct result of the *great recession* that led to a decrease in demand, coupled with the increase in production. At the time of writing, NYMEX natural gas futures are trading at about \$4 per million Btu (about \$4.50/1,000 ft³) and they were as high as \$6 per million Btu (about \$6.80/1,000 ft³) at the beginning of 2010. Note that extremely low price levels are not likely to continue for long if they are below the production costs. Extreme price volatility can be expected when demand and supply (or capacity) go neck to neck.

In this context, one should also mention the problems associated with natural gas transportation via pipelines. In January 2009, the price quarrel between Russia and Ukraine ended in Russia cutting off all gas flows through Ukraine for 13 days. This resulted in severe distress in Southeastern Europe, which is nearly completely reliant on Russian gas, and other parts of Europe. While not an issue in North America, political

stability of producer and transit countries is a significant concern for European countries that can make them reluctant in investing in natural gas-fired power plants. Alternate routes between Russia, Azerbaijan, and other countries, such as thousands of miles long *South Stream* and *Nabucco* pipelines, are actively pursued projects to provide answers to such concerns. The only other practical way to transport natural gas is to move it by sea by cooling it to -162°C (-260°F) and converting it to a liquid (liquefied natural gas or LNG), which reduces its volume by 600 times. Specially built, double-hulled cryogenic ships are used to transport LNG from its port of origin to its port of destination, which are purpose-built facilities used exclusively to export or import LNG. From the receiving port, LNG is shipped to its final destination via tanker trucks, trains, or pipelines (after regasification). Expensive infrastructure investment requirements, safety concerns, large “carbon footprint” during liquefaction, transport, and regasification are key concerns associated with LNG, which limit its share in the overall natural gas consumption (currently less than 10% of US imports).

In terms of cost, regulatory costs, and incentives, e.g., those associated with GHG emission caps and cap-and-trade system requirements (e.g., for manufacturers and utilities to purchase pollution permits), should be considered along with straight dollars per Btu. Coal-fired generation can compete with natural gas in terms of GHG and other pollutant emissions only at a significantly high capital cost and reduced efficiency. Currently, there are two major approaches to accomplish this: Postcombustion carbon capture and storage (CCS) technologies that can clean the stack gas of the existing (via retrofit) and new coal-fired plants and precombustion technologies that are a very good fit with *integrated gasification* CC (IGCC) power plants. Note that carbon emission abatement technologies via stack gas treatment are equally applicable to natural gas-fired CC power plants. In terms of both efficiency and cost, detailed studies indicate that NG-fired CC technology is superior to oil or coal-fired plants as well as IGCC [85].

Objectively viewed, these factors are strong incentives for natural gas burning power plants to replace aging, inefficient, and carbon-emitting, coal-fired power plants in the USA. To a certain extent, this can

be accomplished even without building new natural gas-fired CC power plants. A recent report estimated that CC power plants had an average 2007 capacity factor of 42% [86]. In comparison, in the same year, coal-fired plants had a much higher capacity factor of 75%. Clearly, at least on paper, the replacement potential by simply switching generation from coal-fired to natural gas-fired plants is pretty big. In reality, factors such as geographical proximity of the power plants to each other and to the locations of highest demand, transmission system limitations and dispatch issues limited this potential severely; i.e., to about 5–9% of total US coal-based electricity generation [86]. While the proposed switch from coal to natural gas would also require an increase of 5–20% in natural gas supply, this should not be a significant problem in the light of recent developments discussed in the earlier paragraphs. A recent article states that several major US utilities have announced that they will replace at least some of their coal-fired plants slated for decommissioning (total capacity of more than 4.5 GW) [10].

Another factor that favors the natural gas-fired gas turbine plants over the coal-fired steam turbine power plants is operational flexibility. The rapid start-up and good turndown characteristics of gas turbine simple and combined cycle power plants, especially with the increase of renewable energy power plants (mostly solar and wind) connecting to the grid, make them eminently suitable to cyclic on-demand duties to compensate for the variability of the renewable generation. Presently, all major OEMs provide CC plants that can reach base load (several hundred megawatts) within 30–60 min without sacrificing base load efficiency, parts life, and RAM. Aeroderivative gas turbines, such as GE's intercooled LMS-100, can provide 100 MW within 10 min in an emergency or for peak shaving in hot summer months. Especially in Europe and to a certain extent in the USA, large, highly flexible natural gas-fired CC power plants that start and shut down daily for up to 250 starts and a total of 4,000 h of operation per year are expected to supplement renewable power generation. These plants are based on advanced F-Class and H-Class gas turbines that are rated at 58–59% efficiency and expected to reach more than 60% in the near future.

Clearly, then, no matter how regional politics in the USA, Europe, and other parts of the world play out, and what kind of legislative packages eventually emerge from governing bodies to control, cap or regulate carbon and other GHG emissions, natural gas-fired electricity generation with advanced gas turbine-based power plants will be an important part of the overall generating technology portfolio. In all likelihood, natural gas-fired advanced CC power plants will be the bridge technology between today's coal-based power generation and tomorrow's renewable (and probably nuclear) power generation to alleviate the lurking problem of global warming. (In fact, this is pretty much the conclusion reached by a recent MIT report, which was made public after the writing of the present article [87].) The question is what will be the primary technologies in the next, say, quarter century?

Without a doubt, gas turbine-based power plants, in simple cycle (both heavy-duty industrial and aeroderivative variants) and combined cycle configurations will be by far the dominant technology in natural gas electric power generation. Today's advanced air-cooled F-Class and H-Class, steam-cooled G-Class, GE's H-SystemTM, and J-Class engines with firing temperatures up to 1,600°C and PRs up to 23 (up to 35 in reheat machines) are currently capable of 58%, have recorded 59%, and are slated for 60% (or even 61% as claimed by one OEM) in 2011–2012 [88]. These are large units with 500–600 MW per block in 1×1 configuration designed not only for base load duty but for operational flexibility as described earlier. (In passing, note that the cited efficiencies are typically ISO rating numbers that reflect the most favorable operating assumptions (e.g., once-through open-loop cooling with water drawn from a natural source such as river, ocean, etc). Depending on site and ambient characteristics, performances of actual installed units will be somewhat lower with wide variability.)

Increasing power plant efficiency is the simplest and arguably the easiest way to reduce carbon emissions without breaking the bank or worrying about what to do with all that CO₂ or where to put it. As mentioned earlier, ambitious targets as high as 65% or even 70% have been in circulation for some time [57–60]. The discussion in the preceding sections of the

article hopefully makes it clear that the key to ever-increasing gas turbine combined cycle efficiencies is increasing firing temperatures and commensurate cycle pressure ratios. The factors that prevent easy realization of that simple goal are primarily related to available component materials that can withstand the extreme pressures and temperatures and limits imposed by pollutant emission regulations. State-of-the-art nickel-based alloys used in turbine HGP components require significant amounts of cooling air supplied from extraction ports in the compressor, which is detrimental to performance. Higher flame zone temperatures in DLN combustors result in excessive NO_x and/or CO in flue gas, which is incompatible with increasingly stringent environmental regulations that limit the emissions of those pollutants to low single-digit numbers.

Single-crystal (SC) and directionally solidified (DS) superalloys with TBCs helped push firing temperatures up to 1,500°C. With closed-loop steam cooling and advanced film cooling techniques, J-Class turbines are rated at 1,600°C. Current research efforts focus on advances in materials, e.g., ceramic matrix composites (CMC) [58] or metal foams [59], which can reduce or even eliminate cooling air requirements at extremely high HGP temperatures. Manufacturing techniques have always been a significant driver in the implementation of aforementioned innovations; e.g., TBC application processes (e.g., electron beam physical vapor deposition), casting and machining technologies to produce SC airfoils. In a similar vein, laser-drilling and casting methods to achieve transpiration (effusion) cooling (the ultimate limit of film cooling) are related areas of research and development to further the goal of increasing turbine inlet temperatures. An associated effort is made on the bottoming cycle side, i.e., development of HRSG heat exchanger tubes, steam headers and piping and steam turbine materials (casing, rotor, and buckets) that can withstand increased steam conditions as high as 180 bar and 600–700°C [89].

Closed-looped steam cooling of the combustor transition piece and stage-1 stators (nozzles) is an existing technique to increase the firing temperature without raising the combustor exit (turbine inlet) or flame temperatures [32, 33]. This enables performance improvement without a concomitant increase in NO_x

emissions. Another method to facilitate extreme combustor exit and firing temperatures (up to 1,700°C) is exhaust gas recirculation (EGR), in which a fraction of the stack gas (up to 35%) is recirculated into the compressor or combustor inlet [90]. EGR reduces thermal NO_x production via reduced O₂ in combustion air and peak temperatures in flame. EGR's benefit becomes more pronounced at increased flame temperatures and it can reduce NO_x emissions by up to 50% [91]. EGR is also beneficial for postcombustion CCS because it increases the CO₂ concentration in stack gas (from ~3–4% to ~8–10% by volume), which leads to a reduction in the capital cost of the separation plant [91].

Lean premix or DLN combustion technology is the industry standard for heavy-duty industrial gas turbines for large-scale electricity production. In terms of the ability to meet the stringent emissions requirements, the technology is nearly at its limit at the current turbine inlet temperatures. EGR is one possibility to extend the applicability of DLN or even diffusion combustors to 1,700°C turbine inlet temperatures with low NO_x emissions. Another way to mitigate the NO_x emissions is selective catalytic reduction (SCR) in the HRSG. SCR is an existing postcombustion NO_x reduction technology. Other combustion technologies are being investigated for deployment in IGCC and other low-Btu syngas and blast-furnace gas applications to extend the fuel flexibility of gas turbines. The reader can consult the relevant sections in *The Gas Turbine Handbook* published by NETL (US DOE, Office of Fossil Energy). Various high-hydrogen and oxyfuel combustion technologies are being actively investigated for CO₂-free emissions. Since the subject of the present article is natural gas-burning technologies, the reader is referred to relevant academic journals such as *ASME's Journal of Engineering for Gas Turbines and Power*, which contain a wide selection of recent research papers on the subject.

A potentially promising combustion technology that directly attacks the single largest source of exergy destruction in the GT Brayton cycle is pulse detonation combustion (PDC). Originally proposed for jet engine aircraft propulsion, the technology has been around for more than a half century [92]. Recently, PDC is actively considered for land-based power generation [93]. The

technology involves combustion of a fuel-air mixture across detonation waves in tubular devices [94]. Detonation waves are composite waves comprised of a “frozen” shock wave with pressure and temperature jump followed by a relaxation zone where chemical combustion reactions take place. The benefit of the process is temperature *and* pressure rise during the cycle heat addition process with a correspondingly higher mean-effective temperature vis-à-vis the conventional diffusion or DLN combustors with approximately constant pressure (actually a 5–6% pressure loss). Detailed cycle calculations indicate that a PDC gas turbine in CC configuration is potentially 1.5–3 percentage points advantage over standard units at comparable firing temperatures and pressure ratios [95].

SOFC- or MCFC-based hybrid fuel cell plus gas turbine power plants are by far the most efficient future technologies to “burn” natural gas for electric power generation. In these machines, the fuel cell replaces the combustor to oxidize the gaseous fuel. Different configurations are possible. Detailed studies show that fuel cell-GT hybrid plants are capable of thermal efficiencies approaching 70%. Bhargava et al. [96] can be consulted for more information on these and other advanced GT-based power plants such as humid air turbine (HAT) and intercooled recuperated engine (ICR). Almost all variants have been studied in detail and reached various stages of development and commercialization in the past. They have not been able to pass the audition in terms of reliability, operability, capital cost, scalability, and other important considerations (e.g., water consumption) to present viable alternatives to simple Brayton cycle GT in CC configuration. Closed-loop steam cooling and reheat (also known as *sequential combustion*) have been the only successful and commercially accepted variations on the basic cycle. At this time, to the best of the authors' knowledge, there is no reason to expect this situation to change in the foreseeable future; at least in terms of large-scale electric power generation. As always, dual-fuel diesel engines capable of burning gaseous fuels will continue to play a role in small-scale and/or distributed generation applications. The same can be true for very small (several kilowatts) fuel cell-based hybrid units.

Bibliography

Primary Literature

1. Energy Information Administration (EIA) (2009) Annual energy review 2008. <http://www.eia.doe.gov/aer>. Accessed 29 June 2009
2. Energy Information Administration (EIA) (2008) International energy outlook 2008. DOE/EIA-0484(2008). www.eia.doe.gov/oiaf/ieo/index.html.
3. United States Environmental Protection Agency (EPA) (2010) Methane. www.epa.gov/methane.
4. Liss WH, Thrasher WR (1992) Variability of natural gas composition in select major metropolitan areas of the United States. Gas Technology Institute, Chicago, GRI-92/013
5. Natural Gas and the Environment (2010) From <http://www.naturalgas.org/environment/naturalgas.asp>
6. Turbomachinery International (2008) 49(6), 10/2008, Handbook 2009. www.turbomachinerymag.com
7. Wärtsilä 50DF generating set (2010) From www.wartsila.com
8. Khan BH (2006) Non-conventional energy sources. Tata McGraw Hill, New Delhi, India
9. McNeely M (2006) Power generation order survey. Diesel & Gas Turbine Worldwide. Article from Oct 2006 issue
10. Burt B, Mullins S (2010) U.S. gas-fired power development: last man standing, POWER. <http://www.powermag.com>. Accessed Sept 2010, pp 71–73
11. Wilson DG, Korakianitis T (1998) The design of high efficiency turbomachinery and gas turbines, 2nd edn. Prentice-Hall, Uppersaddle River
12. Von Ohain H (1996) Foreword. In: Mattingly JD (ed) Elements of gas turbine propulsion. Tata McGraw Hill, New Delhi, India, Edition 2005
13. Meher-Homji CB (1997) The development of the Junkers Jumo 004B – the world's first production turbojet. J Eng Gas Turb Power 119:783
14. Meher-Homji CB (1998) The development of the Whittle Turbojet. J Eng Gas Turb Power 120:249
15. Meher-Homji CB (2000) Pioneering turbojet developments of Dr. Hans von Ohain – from the HeS 1 to the HES 011. J Eng Gas Turb Power 122:191
16. Soares C (2006) Gas turbines in simple cycle and combined cycle applications. Section 1.1 in The gas turbine handbook. US DOE, Office of Fossil Energy, NETL. <http://www.netl.doe.gov/technologies/coalpower/turbines/refshelf/handbook/TableofContents.html>
17. Leiste V (1999) Development of the siemens gas turbine and technology highlights. Siemens Power Generation, Erlangen, Germany
18. Miller H, Nemec T (2006) Gas turbines. In: Myer K (ed) Mechanical engineers' handbook, 3rd edn., Energy and power. Wiley, Hoboken, Chapter 24
19. Brandt D (2007) A brief history of GE energy product lines. General Electric Company, New York
20. Brandt D (1988) The design and development of an advanced heavy-duty gas turbine. J Eng Gas Turb Power 110:243–250
21. Stodola A (1927) Steam & gas turbines. Authorized translation from the 6th German edition: Löwenstein LC. McGraw-Hill, New York
22. Langston LS (2010) World's first gas turbine power plant. ASME Mech Eng 132(4):51
23. Tomlinson LO, Lee DT (1985) Combined cycles. In: Sawyer JW, Japikse D (eds) Sawyer's gas turbine engineering handbook, Chapter 7. Turbomachinery International Publications, Norwalk, Conn., USA
24. Horlock JH (1994) Combined cycle power plants – past, present, and future. J Eng Gas Turb Power 117:608–616
25. Gebhardt E (2000) The F Technology experience story, GER-3950C. <http://www.gepower.com/>
26. Haselbacher H (1989) Gas turbine fundamentals. In: Elliott TC (ed) Standard handbook of power plant engineering. New York, McGraw-Hill Publishing POWER magazine
27. Horlock JH (1997) Aero-engine derivative gas turbines for power generation: thermodynamic and economic perspectives. J Eng Gas Turb Power 119:119–123
28. Moran MJ, Shapiro HN (1988) Fundamentals of engineering thermodynamics. Wiley, New York
29. Cohen H, Rogers GFC, Saravanamuttoo HIH (1987) Gas turbine theory, 3rd edn. Longman, London
30. Cumpsty N (2003) Jet propulsion, 2nd edn. Cambridge University Press, Cambridge
31. Schilke PW (2004) Advanced gas turbine materials and coatings, GER-3569 G. www.gepower.com
32. Pritchard JE (2003) H-System™ technology update. GT2003-38711, ASME turbo expo – power for land, sea & air, 16–19 June 2003, Atlanta
33. Koenke C (2006) Steam cooling of large frame gas turbines one decade in operation. VDI Ber Nr 1965:33–42
34. Imwinkelried B (1995) Advanced cycle system gas turbines GT24/GT26: the highly efficient gas turbines for power generation. In: Proceedings of the 21st international congress on combustion engines, CIMAC 1995, Interlaken, Switzerland
35. Chiesa P, Macchi E (2004) A thermodynamic analysis of different options to break 60% electric efficiency in CC power plants. J Eng Gas Turb Power 126:770–785
36. Holland MJ, Thake TF (1980) Rotor blade cooling in high pressure turbines. J Aircr 17:412–418
37. Elmasri MA, Pourkey F (1986) Prediction of cooling flow requirements for advanced utility gas turbines. Part 1: analysis and scaling of the effectiveness curve, 86-WA/HT-43, ASME Winter Annual Meeting, Anaheim, 7–12 Dec 1986
38. Elmasri MA (1986) Prediction of cooling flow requirements for advanced utility gas turbines. Part 2: influence of ceramic thermal barrier coatings. ASME Winter Annual Meeting, Anaheim, 7–12 Dec 1986
39. Elmasri MA (1985) On thermodynamics of gas turbine cycles: part 1 – second law analysis of combined cycles. J Eng Gas Turb Power 107:880–889

40. Elmasri MA (1986) On thermodynamics of gas turbine cycles: part 2 – a model for expansion in cooled turbines. *J Eng Gas Turb Power* 108:151–159
41. Elmasri MA (1986) On thermodynamics of gas turbine cycles: part 3 – thermodynamic potential and limitations of cooled reheat gas turbine combined cycles. *J Eng Gas Turb Power* 108:160–170
42. Horlock JH, Watson DT, Jones TV (2001) Limitations on gas turbine performance imposed by large turbine cooling flows. *J Eng Gas Turb Power* 123:487–494
43. Horlock JH (2001) The basic thermodynamics of turbine cooling. *J Eng Gas Turb Power* 123:583–591
44. Wilcock RC, Young JB, Horlock JH (2005) The effect of turbine blade cooling on the cycle efficiency of gas turbine power cycles. *J Eng Gas Turb Power* 127:109–120
45. Young JB, Wilcock RC (2002) Modeling the air-cooled gas turbine: parts 1 and 2. *J Turbomach* 124:207–222
46. Gülen SC (2010) A simple mathematical model for cooled gas turbines. GT2010-22160, ASME turbo expo – power for land, sea & air, 14–18 June 2010, Glasgow
47. Rice IG (1995) Steam-injected gas turbine analysis: steam rates. *J Eng Gas Turb Power* 117:347–353
48. Cheng DY, Nelson ALC (2002) The chronological development of the Cheng cycle steam injected gas turbine during the past 25 years. ASME International – IGTI Turbo Expo 2002, GT2002-30119
49. Rao A (1989) Process for producing power. U.S. Patent No. 4,289,763
50. Adelman ST, Hoffman MA, Baughn JW (1995) A methane-steam reformer for a basic chemically recuperated gas turbine. *J Eng Gas Turb Power* 117:16–23
51. McDonald CF, Boland CR (1981) The Nuclear Closed-Cycle Gas Turbine (HTGR-GT) – dry cooled commercial power plant studies. *J Eng Gas Turb Power* 103:89–100
52. Reale MJ (2004) New high efficiency simple cycle gas turbine – GE's LMS100™, GER-4222A. www.gepower.com
53. Mercury 50, Recuperated Gas Turbine Generator Set, Solar® Turbines (2010) www.solarturbines.com
54. Cox JC, Hutchinson D, Oswald JI (1995) The Westinghouse/Rolls Royce WR-21 gas turbine variable area power turbine design. ASME Paper 95-GT-54, International Gas Turbine and Aeroengine Congress and Exposition, Houston, TX, 5–8 June 1995
55. Hofer DC, Gülen SC (2006) Efficiency entitlement for bottoming cycles. GT2006-91213, ASME turbo expo – power for land, sea & air, Barcelona, Spain, 8–11 May 2006
56. Gülen SC, Smith RW (2008) Second law efficiency of the Rankine bottoming cycle of a combined cycle power plant. ASME Paper GT2008-51381. ASME turbo expo 2008, Berlin, Germany, 9–13 June 2008
57. Bohn D (2006) SFB 561: Aiming for 65% CC efficiency with an air-cooled GT, Modern Power Systems, pp 26–29, Sept 2006
58. Mutassim Z (2008) New gas turbine materials. *Turbomachinery International*, Sept/Oct 2008 issue, 38–42
59. Bohn D, Diltthey U, Schubert F (2004) Innovative Technologien für ein GuD-Kraftwerk mit 65% Wirkungsgrad. VDI-Berichte Nr 1857:13–25
60. Rao AD, Robson FL, Geisbrecht RA (2002) Power plant system configurations for the 21st century. In: ASME turbo expo 2002, Amsterdam, the Netherlands, 3–7 June 2002
61. Lundberg WL, Veyo SE, Moeckel MD (2003) A high efficiency solid oxide fuel cell hybrid power system using the mercury 50 advanced turbine system gas turbine. ASME J Eng Gas Turb Power 125:51–58
62. Massardo AF, Lubelli F (2000) Internal Reforming Solid Oxide Fuel Cell – Gas Turbine Combine Cycles (IRSOFC-GT); Part I: cell model and cycle thermodynamic analysis. ASME J Eng Gas Turb Power 122:27–35
63. Massardo AF, Magistri L (2003) Internal Reforming Solid Oxide Fuel Cell – Gas Turbine Combine Cycles (IRSOFC-GT); Part II: exergy and thermoeconomic analyses. ASME J Eng Gas Turb Power 125:67–74
64. Chase DL, Kehoe PT. GE combined-cycle product line and performance, GER-3574 g. GE Energy
65. Chris EM, Leroy OT. GE combined-cycle experience, GER-3651. <http://www.gepower.com>.
66. Tomlinson LO, McCullough S. Single-shaft combined – cycle power generation system, GER-3767c. <http://www.gepower.com>
67. Matta RK, Mercer GD, Tuthill RS. Power systems for the 21st century – H GT combined-cycles, GER-3935B. GE Energy
68. Smith RW, Polukort P, Maslak CE, Jones CM, Gardiner BD. Advanced technology combined cycles, GER-3936a. GE Power Systems
69. Phylipsen GJM, Blok K, Worrell E (1998) Handbook on international comparisons of energy efficiency in the manufacturing industry. Department of Science, Technology and Society, Utrecht University, The Netherlands
70. Gülen SC (2010) A proposed definition of CHP efficiency, POWER. <http://www.powermag.com>, pp 58–63, June 2010
71. European Association for the Promotion of Cogeneration (Mar 2001) A guide to cogeneration. http://www.cogeneurope.eu/wp-content/uploads/2009/02/educogen_cogen_guide.pdf
72. Energy Information Administration (EIA) (2009) Annual energy review 2008. <http://www.eia.doe.gov/aer>. Accessed 29 June 2009
73. Energy Information Administration (EIA) (2010) Electric power annual. <http://www.eia.doe.gov/fuelelectric.html>. Accessed 20 Jan 2010
74. Davis LB, Black SH (2000) Dry low nox combustion systems for GE heavy-duty gas turbines, GER-3568 g. <http://www.gepower.com>
75. Lefebvre AH (1995) The role of fuel preparation in low-emission combustion. *J Eng Gas Turb Power* 117:617
76. Roointon, Pavri, Moore, Gerald D (2001) Gas turbine emissions and control, GER-4211. <http://www.gepower.com>

77. Hilt MB, Waslo J (October 1984) Evolution of NO_x abatement techniques through combustor design for heavy-duty gas turbines. *J Eng Gas Turb Power* 106:825
78. Touchton GL (1984) An experimentally verified NO_x prediction algorithm incorporating the effects of steam injection. *J Eng Gas Turb Power* 106:833
79. Davi MA (1994) GE gas turbine combustion flexibility, GER-3946. GE Energy
80. Miller HE (1994) Development of the GE quiet combustor and other design changes to benefit quality, GER-3551. <http://www.gepower.com>
81. Peters M, Timmerhaus K, West R (2004) Plant design and economics for chemical engineers, 5th edn. McGraw-Hill, London
82. Bejan A, Tsatsaronis G, Moran M (1996) Thermal design & optimization. Wiley, New York
83. Kehlhofer R, Warner J, Nielsen H, Bachmann R (1999) Combined cycle gas & steam turbine power plants, 2nd edn. PennWell Corp, Tulsa
84. As reported in the press per Potential Gas Committee report (2008) Potential supply of natural gas in the United States, Potential Gas Agency, Colorado School of Mines, Golden, 31 Dec 2008
85. Gambini M, Vellini M (Jan 2003) CO₂ emission abatement from fossil fuel power plants by exhaust gas treatment. *J Eng Gas Turb Power* 125:365–373
86. Wagman D (2010) Can natural gas displace coal? *Power Eng* (Mar 2010 issue): 4
87. The future of natural gas – an interdisciplinary MIT study (2010) Interim report by MIT Energy Initiative, ISBN 978-0-9828008-0-5, Massachusetts Institute of Technology, Boston
88. Robb D (2010) CCGT: breaking the 60 percent efficiency barrier. *Power Eng Int* 18(3). www.peimagazine.com
89. Review of status of advanced materials for power generation, Technology Status Report, Cleaner Coal Technology Programme, Department of Trade and Industry (Oct 2002) London
90. Tukagoshi K, Muyama A, Uchida S et al (Oct 2005) Latest technology for large capacity gas turbine. *MHI Tech Rev* 42(3)
91. ElKady AM, Evulet A, Brand A (May 2009) Application of exhaust gas recirculation in a DLN F-class combustion system for postcombustion carbon capture. *J Eng Gas Turb Power* 131:#034505
92. Kailasanath K (2000) Review of propulsion applications of detonation waves. *AIAA J* 38(9):1698–1708
93. Goldmeer J, Tangirala V, Dean A (2008) System-level performance estimation of a pulse detonation based hybrid engine. *J Eng Gas Turb Power* 130:#011201
94. Tangirala VE, Rasheed A, Dean AJ (2007) Performance of a pulse detonation combustor-based hybrid engine, GT2007-28056. ASME turbo expo – power for land, sea & air, Montreal, Canada, 14–18 June 2007
95. Gülen SC (2010) Gas turbine with constant volume heat addition. ESDA2010-24817. ASME 2010 10th biennial conference on engineering systems design and analysis, Istanbul, Turkey, 12–14 July 2010
96. Bhargava R, Bianchi M, Campanari S et al (2010) A parametric thermodynamic evaluation of high performance gas turbine based power cycles. *J Eng Gas Turb Power* 132:#022001

Books and Reviews

- Bejan A (2006) Advanced engineering thermodynamics, 3rd edn. Wiley, New Jersey
- Boss M (1996) Steam turbines for STAG™ combined cycle power systems, GER-3582E. <http://www.gepower.com>
- Boyce MP (2006) Gas turbine engineering handbook, 3rd edn. Gulf Professional Publishing, Houston
- Chase D (2001) Combined cycle development evolution and future, GER-4206. <http://www.gepower.com>
- Colegrove D, Mason P, Retzlaff K, Cornell D (2001) Structured steam turbines for the combined cycle market, GER-4201. <http://www.gepower.com>
- Constant EW II (1980) The origins of the turbojet revolution. The Johns Hopkins University Press, Baltimore/London
- Cotton KC (1998) Evaluating and improving steam turbine performance, 2nd edn. Cotton Fact Inc, Rexford
- Denton JD (1993) Loss mechanisms in turbomachines, The 1993 IGTI scholar lecture. *J Turbomachinery* 115:621–656
- Dunn MG (2001) Convective heat transfer and aerodynamics in axial flow turbines. *J Eng Gas Turb Power* 123:637–686
- Elmasri MA (2007) Design of gas turbine combined cycle and cogeneration systems – theory, practice and optimization. Seminar Notes, Thermoflow, Inc., Sudbury, MA. info@thermoflow.com
- Han JC, Dutta S, Ekkad SV (2000) Gas turbine heat transfer and cooling technology. Taylor & Francis, New York
- Horlock JH (2001) Combined power plants: including Combined Cycle Gas Turbine (CCGT) Plants. Krieger Publishing Company, Malabar
- Kehlhofer R, Hannemann F, Stirnimann F, Rukes B (2009) Combined cycle gas & steam turbine power plants, 3rd edn. PennWell Corp, Tulsa
- Lakshminarayana B (1996) Fluid dynamics and heat transfer of turbomachinery. Wiley, New York
- Lefebvre AH, Ballal DR (2010) Gas turbine combustion: alternative fuels and emissions, 3rd edn. CRC Press/Taylor & Francis, Boca Raton
- Nag PK (2006) Power plant engineering, 2nd edn. Tata McGraw-Hill, New Delhi, India
- Saravanamuttoo HH, Rogers GFC, Cohen H, Straznicky PV (2009) Gas turbine theory, 6th edn. Pearson Prentice Hall, England
- Traupel W (1977) Thermische Turbomaschinen, Erster Band, Thermodynamisch-strömungstechnische Berechnung, 3, neuarbeitete und erweiterte Auflage. Springer, Berlin/Heidelberg/New York

Natural Resource Flows and Sustainability in Urban Areas

STEFAN ANDERBERG

Lund University Centre for Sustainability Studies,
Lund University, Lund, Sweden

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Urban Sustainability Challenges
Urban Planning
Cities as Ecosystems
Urban Metabolism
Future Directions
Bibliography

Glossary

Ecological footprint A measure of the resource consumption of a given unit (e.g., a city and its population) that represents the area of ecologically productive land and water needed for the production and assimilation of wastes generated.

Emergy analysis Emergy is the amount of energy, usually solar, that is directly or indirectly required to generate a given output flow. Emergy analysis allows analysis of the whole embodied environmental work expressed in a single unit and has been applied to flows in cities and other regions as well as different types of products and production processes.

Urban ecology Research on urban ecosystems and applications of ecological principles in connection with urban planning.

Urban metabolism Involves conceptualizing a city as an organism or ecosystem and tracking resources that go into the system and products and wastes that leave it.

Urban mining Recovery of metals and other valuable materials in urban areas. Studies that have identified metal concentrations comparable to mines in cities and urban waste dumps suggested that

an increasing share of the raw materials for metal production in the future will be recovered from wastes in urban areas.

Definition of the Subject and Its Importance

Cities and urban areas are of central importance for global sustainable development both as dominating sites of production and consumption and as contexts for developing new, more sustainable practices. All cities depend on large imports of energy and other natural resources to satisfy consumption of their inhabitants as well as local production, trade, and services. These resource flows are closely linked to global and regional sustainability issues, such as resource scarcity, pollution, and competition for land or water, as well as local health, environmental, and distributional issues. How resources are consumed and managed in the expanding cities has strong implications for the global resource flows and related pressures in different scales. The goal of “sustainable development” puts into question many traits of current urbanization and city development and calls for a sustainable urban transformation. Such a transformation needs to rely on sound, efficient, and sustainable resource management and sustainable urban structures in terms of the built environment, transport systems, and green and blue structures. The handling of resource flows in cities and urban areas is the core of this challenge. For increasing the capacity of local strategic action toward a more sustainable use of resources, it is essential to develop improved understanding of the urban resource flows. Studies of urban ecosystems and resource flows as well development of ecological planning approaches connect to this challenge.

Introduction

After centuries of intense urbanization, more than half of the world's population lives in urban areas. The global urban population is expected to grow by more than 50%, or 1.7 billion people, between 2008 and 2030 [1]. Therefore, the cities are increasingly recognized as the most important global sustainability challenge of this century [2]. The urban ecological sustainability challenges are both linked to global resource scarcity, pollution and biodiversity issues,

and local health and environmental problems as well as resource access and distribution among different groups of people. With their central role in connection with manufacturing, trade, and consumption, big cities since long dominate global energy use and material flows [3]. How resource use develops in cities have strong implications for resource flows in different scales and related pressures. How cities are built, heated and cooled; how efficient their infrastructures are, and what urban citizens consume or how they travel in their everyday lives have great influence on greenhouse emissions, and exploitation of scarce land, water and mineral resources in different parts of the world. For increasing the capacity of local action toward a more sustainable use of resources, it is essential to develop improved analysis tools that allow a better guidance for strategic action [4, 5]. A key challenge is to improve the overview and understanding of the environmental linkages of the urban resource flows in different scales and develop urban governance that contributes to a more sustainable use of resources [6, 7]. Studies of urban ecosystems and resource flows as well as development of sustainability planning connect to this challenge. Urban metabolism studies explore the interactions among resource flows, urban transformation processes, waste streams, and quality of life and connect to the challenge of combining urban planning with sustainable resource management. This entry focuses particularly on natural resource flows in urban areas and urban metabolism studies, their results and challenges, and their potential contributions to sustainable urban planning and governance.

Urban Sustainability Challenges

Cities alter their local environments and influence the global environment in several ways. They occupy and modify space and are characterized by intense competition for space between different uses of land [8]. They import vast quantities of food, water, and energy and export emissions and waste, and they depend in various ways on local and global ecological services. Cities have always been a scene for unsafe living conditions and ecological problems [9]. Historically, it is possible to distinguish three different but often overlapping types of urban sustainability crises during the industrial era:

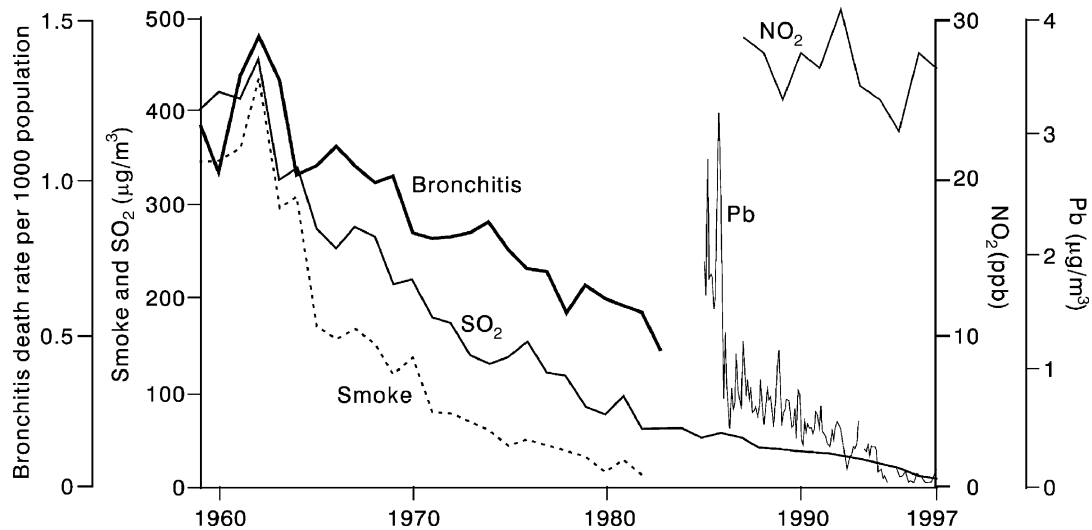
- A *hygienic* crisis connected to unsafe drinking and inadequate sanitation and waste management, and related epidemics in fast growing cities
- A *social* crisis connected to poverty; insufficient, poor, and unhealthy housing; and lacking access to light, fresh air, and green spaces
- An *environmental* crisis particularly connected to severe pollution and health hazards for local people

In cities in industrialized regions, these crises have generally been successfully addressed via development of urban infrastructures, planning, organization, and control, and the most urgent local problems have disappeared [10]. Local pollution problems still remain in most cities, mostly connected to traffic, waste, and diffuse pollution. Figure 1 shows the development in Manchester as an example of the improvement of environmental quality and health in European cities toward the end of the twentieth century. By the end of the nineteenth century, the bronchitis death rate was above 3 per 1,000 inhabitants compared with less than 0.5 in the 1980s.

In fast-growing cities in the developing world, these three types of problems still remain, overlap, and often get worse. In many cities in Africa, Asia, and Latin America, large parts of the inhabitants live under unhealthy conditions in poor housing without access to safe drinking water, sanitation, or other basic services [11, 12]. In great contrast to the desperate needs for basic infrastructure in Third World cities, the emerging sustainable city efforts in Europe, North America, and Australia during recent decades have, to a large extent, been concerned with the unsustainability of the cities in a global context. Global change problems, such as resource scarcity, climate change, land use competition, or biodiversity, put into question the energy use, spatial structures, and infrastructural arrangements of modern society in general and big city regions in particular. In this context, promoting efficient and sustainable management of resource flows in the city has become of central importance.

Urban Planning

Societal development is in different ways reflected in changed flow and flow patterns. Industrialization, increasing wealth, expansion of the city, changed design of buildings and city areas, and changed



Natural Resource Flows and Sustainability in Urban Areas. Figure 1

Development of air pollution (smoke, SO_2 , lead, and NO_x) and bronchitis death rate (per 1,000 inhabitants) in Manchester (Source: [10])

patterns of settlement, work, commuting, production, consumption, trade, leisure activities, and use of the urban landscape of the big city all result in changed flows. Resource flows are thus strongly influenced by urban planning and design of urban areas. Urban planning has developed in interaction with different social welfare issues, e.g., local environmental problems. It has traditionally primarily focused on regulating the location and intensity of activities to avoid health problems and environmental degradation and influenced resource flows, but planning has seldom comprehensively and explicitly addressed resource management [13]. Urban planning emerged as a response to the unhealthy and polluted conditions in the fast-growing industrializing European cities in the nineteenth century. The overwhelming problems were gradually overcome by investments in infrastructure and introduction of a system of laws and authorities to take care of waste collection, water, hygienic control, and worker protection. Society took on a growing responsibility for the health and well-being of its citizens, and the major tools have been infrastructural investments and an increasingly ambitious spatial planning. During the twentieth century, the social crisis connected to poor, unhealthy housing conditions was, in European cities, addressed by demolition, renovation and social housing projects, and establishment of

parks, green and recreational areas, and separation of industry and housing. Cities were, in the twentieth century, transformed by motorized road traffic and the breakthrough of the private car, but also visions of more human and ecologically sound urban environments influenced the urban development. Both decentralized city visions, such as Howard's "Garden City," and more centralized visions, such as Le Corbusier's "Radiant City," have influenced urban and suburban developments.

With the growing industrial production, increasing use of fossil fuels, cities in general were severely polluted until the mid-twentieth century. From the 1960s to the 1990s, increasing concern for the environment stimulated important infrastructural investments in wastewater treatment, district heating, energy source substitution, which, together with improved pollution control and industrial restructuring, decreased the industrial pollution and health hazards for urban inhabitants in Europe and North America.

Planning for Urban Sustainability

From the 1970s, growing concern about environment and resource use also influenced urban planning. New urban forms, such as the compact city with

high-density settlement and mixed uses, were promoted as a response to urban sprawl and high dependency on cars and fossil fuel consumption after the oil crisis. The real breakthrough for environmental and ecological planning and interest for the relationship between planning and resource management came with the goal of “sustainable development” that became influential from the late 1980s when ecocity or sustainable city visions and ambitions spread rapidly through national, international, and intercity initiatives. UN Habitat, the World Bank, and the EU have launched sustainable urban development programs that, in different ways, support cities in their strive for a more sustainable development. In Europe, more than 2,500 local and regional governments have signed the Aalborg Charter from 1994 and joined the European Sustainable Cities and Towns Campaign. The most ambitious of these local communities have signed the Aalborg Commitments from 2004.

Sustainable urban development ambitions have brought both broader concerns and new ambitions to urban governance and planning. For increasing sustainability, waste and pollution must be minimized, natural resources conserved, and the carrying capacity of ecosystems respected. The sustainable city agenda includes often compact, efficient land use; reduced car use, efficient public transport, efficient resource use, and reduced pollution and waste; restoration of natural systems; good housing and living environments, community participation and involvement, social equity and inclusion, and a sustainable economy [14]. Sustainability planning requires a holistic view of a city or region that includes equal concern for environmental, economic, and social sustainability [15]. It is conceived as a process, and sustainability, as a goal to work toward rather than something soon to be achieved. In order to achieve such goals, comprehensive plans, zoning ordinances, building codes, and planning processes must be revised for incorporating sustainability principles into urban environmental planning [15–17]. This does not only imply that environmental quality or resource conservation goals must be as important as economic but also that other goals, such as social justice and equity, must be addressed in relation to environmental efforts. For realizing such integrative ambitions, it is often stressed that planning and decision processes must become more open and

integrated: all levels of government, business, civil society, and individual must work together [17].

Sustainability is a long-term goal, which can only be achieved through consistent strategic action over the long term. Sustainable urban development needs to rely on sound resource management and sustainable urban structures. To realize this and to reduce the demand of resources and flows of energy, water, waste, and materials, it is necessary to develop a more effective, integrated, and flexible infrastructure for water supply, waste, sanitation, heating, electricity, and fuels that is resilient to changes in population, economy, and climate and the hydrological regime. Urban systems should be multifunctional, serving more than one technical purpose and be able to integrate ecological, recreational, and aesthetic values. For increasing the capacity of municipalities to act strategically, an important challenge is to develop systems-based analysis tools in collaboration with stakeholders that allow a better overview and guidance for strategic action [4]. This is the challenge that has inspired many studies of urban ecosystems and urban metabolism.

Cities as Ecosystems

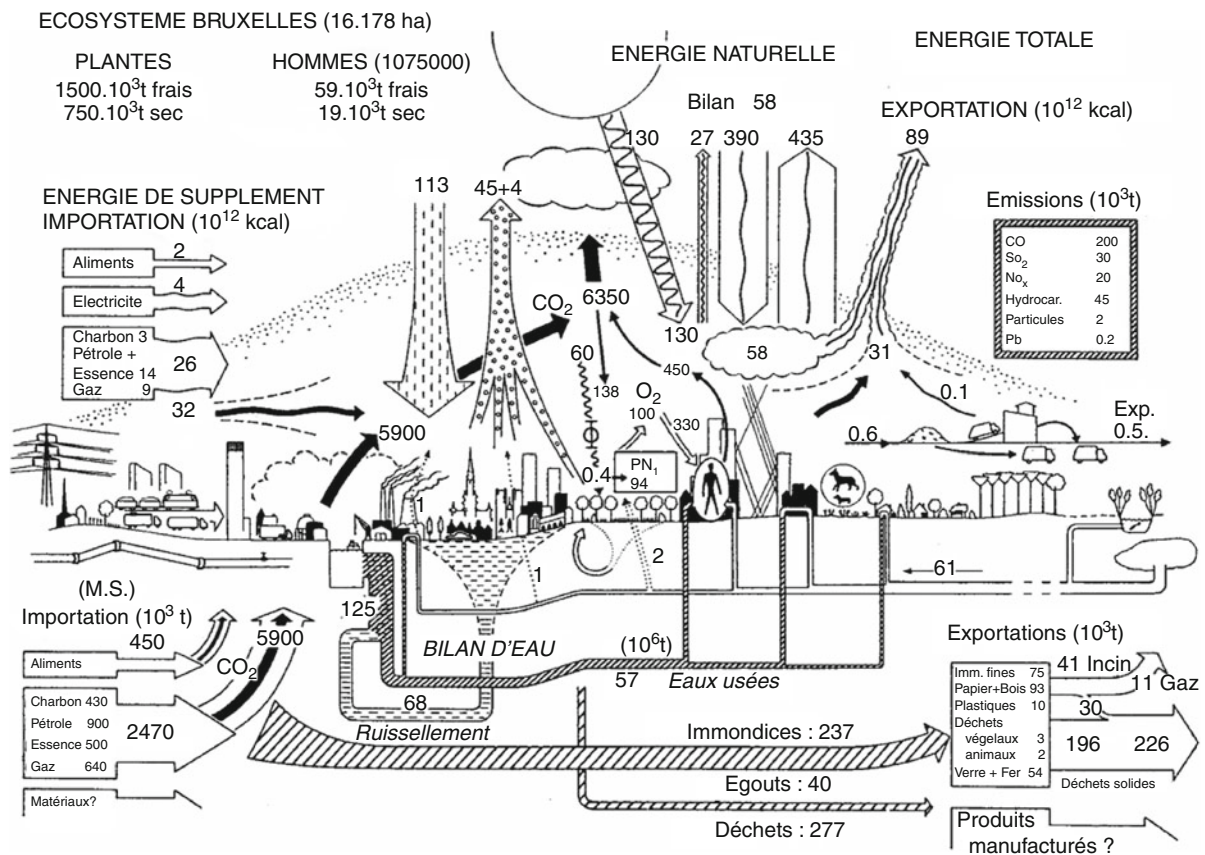
The concept of “urban ecosystems” has proved useful for investigating the relationship between urban areas and the environment [18]. Urban ecosystems are defined by Pickett [19] as “those in which people live at high densities, or where the built infrastructure covers a large proportion of the land surface.” Research on urban ecosystems, often called urban ecology research, has grown and broadened during recent decades. This research can be divided into research on *ecological systems in the city* and on *the city as an ecological system* [20]. The former is dominating and includes many specialized study areas that focus on particular ecological aspects in urban areas [19], e.g., urban climates, soils, hydrology, vegetation, animals, nutrient and carbon flows, as well as different human and social urban systems. The latter analyzes the city as a whole ecosystem based on different ecosystem approaches, human ecosystem or sustainability assessment frameworks or watersheds as integrative tools.

Natural resource flows and their transformations are essential for the functional structure and dynamics

of both natural and human ecosystems. From such a flow-based systems perspective, cities can be described as large stocks of accumulated materials stored in buildings and infrastructure, and they are characterized by large and concentrated flows of raw materials, energy, goods, waste, people, and information. In thermodynamic terms, cities are open systems that depend on imported energy and matter. The cities require vast amounts of resources to function. In the past, the amount of accessible resources in the nearest hinterland was a constraint to the growth of cities [21], but with improved transport and development of international trade, cities no longer rely on their immediate hinterlands for a growing share of resources, including food. Most environmental problems are connected to these natural resource flows.

Urban Metabolism

Urban flows and resource management in different cities have been analyzed in various studies since the 1970s. Figure 2 shows the urban metabolism of Brussels from the early 1970s from the study of Duvigneaud and Denayeyer-De Smet [22]. The term “urban metabolism” was launched by Abel Wolman [23] in his study on generalized flows in a “typical” city in the USA. Kennedy et al. [24] defines “urban metabolism” as “the sum total of the technical and socio-economic processes that occur in cities, resulting in growth, production of energy, and elimination of waste.” The notion of “metabolism” is based upon analogy with the metabolism of organisms, which is one example of parallels between cities and ecosystems, which are



Natural Resource Flows and Sustainability in Urban Areas. Figure 2

The urban metabolism of Brussels, Belgium, in the early 1970s (From: [22])

central in urban ecology. The term “metabolism” has also become established in connection with material and substance flow analysis in other kinds of man-dominated systems, e.g., “industrial metabolism” studies [25] have analyzed material and substance flows in different regions.

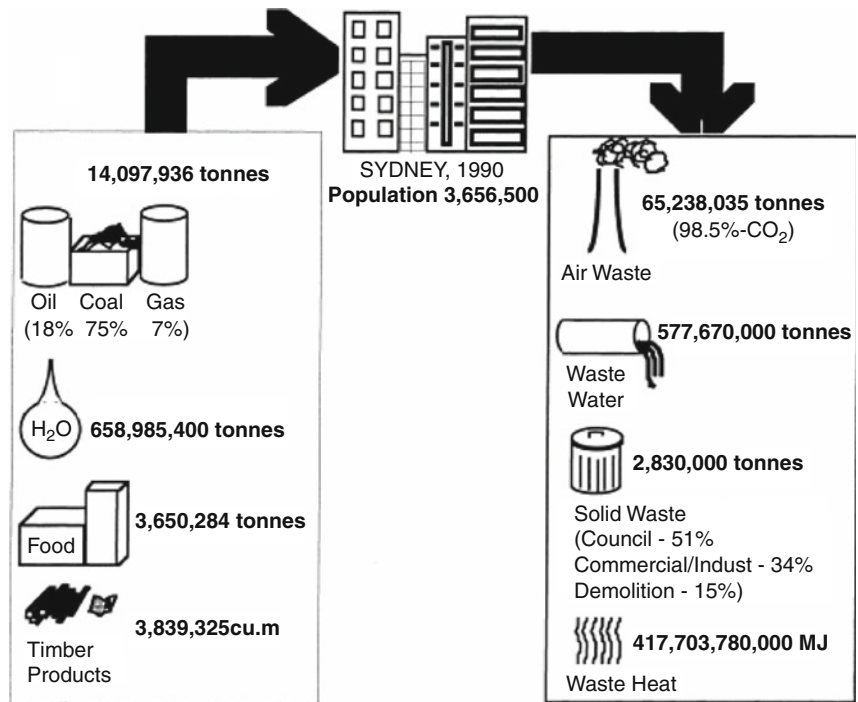
The study of the urban metabolism involves in practice a holistic conceptualization of a city as a system of flows and inputs, outputs, and storage of focused flows, e.g., energy, water, nutrients, materials, and wastes that are then quantified for a certain time period. Figure 3 shows an example of a summary of results in terms of inflows and outflows in the metabolism of Sydney in 1990. By looking at the city as a whole system and by analyzing the pathways along which energy and materials, including pollutants, move, the goal is to improve resource management systems, allow assessment of impacts, and identify technological and organizational opportunities to

increase the efficiency of resource use through, e.g., recycling of wastes as valuable materials and energy conservation [26].

Different Urban Metabolism Studies

Kennedy et al. [27] identified almost 50 “urban metabolism” studies on more than 30 different cities since the 1970s. After a few forerunners in the 1970s, urban metabolism studies reappeared in the 1990s, and the great majority of studies have been published since the year 2000. Hong Kong, Vienna, Stockholm, Toronto, London, and Paris have been subject for several studies.

Most “urban metabolism” studies have focused on energy, water, food (nutrients), materials (metals), and resulting pollution (typically CO₂, SO₂, NO_x, nutrient releases, or heavy metals) and waste in a limited time period. There are also examples of analyses of the historical development, such as for heavy metals in



Notes:
Waste water data do not include stormwater
Timber products and food data derived from national per capita data

Natural Resource Flows and Sustainability in Urban Areas. Figure 3
Resource inputs and waste outputs in Sydney, 1990 (From: [7])

Stockholm [28] and nitrogen in food metabolism in Paris [29] and total material flows in Limerick [30]. In some other cities, such as Hong Kong and Toronto, repeated studies have provided some basis for comparison over time [27].

Most studies have been based on mass or energy measurement for different materials and energy flows, respectively. But some studies have sought to use agglomerated measurement units. Several of the earliest studies in the 1970s e.g., [31] described the metabolism in terms of emergy, an approach which more recently has been applied in connection with analyses of Taipei [32], Beijing [33], and Rome [34]. Related to the urban metabolism concept is also the application of the ecological footprint technique to cities [35]. The studies of Vancouver [36], Santiago [37], and cities of the Baltic region of Europe [38] in the 1990s have been followed by many studies. It has also become increasingly popular to describe the urban metabolism in terms of ecological footprints e.g., [39–42].

Many urban metabolism studies have focused primarily on urban material stocks and flows, e.g., of Lisbon [43], Singapore [44], and York [39]. In connection with heavy metal analyses, the urban stocks have been recognized as future environmental risks or potential raw material sources [45–48]. There has also been a growing interest for food flows and related nutrient flows [49–52]. Urban water metabolism has been addressed in several studies [53–57]. In recent years, urban CO₂ emissions in different cities have received growing interest, and many estimations are built on urban metabolism approaches [58]. There are also many broader studies that in different ways connect to urban metabolism, e.g., transportation of materials [59], and urban infrastructure systems [60]. An indication for the renewed and increasing interest in urban metabolism studies is that two projects have recently started under the EU Seventh Framework Programme: SUME [61] and BRIDGE [62].

Contributions by Urban Metabolism Studies

The urban metabolism studies have contributed to the overview and understanding of the flows and stocks of various materials in urban areas, and of pollution risks. They confirm the importance of cities for energy and material flows in different scales. Furthermore, they

have shown that large parts of the resource flows accumulate in buildings, wastes, and pollution in urban areas. In some cases, they have indicated that accumulation results in surprisingly high concentrations of hazardous materials “hibernating” in cities, which bring future risks and challenges in connection with demolition and waste, diffuse pollution, and pollution of groundwater (e.g., [28]). Brunner and Rechberger [47] have inspired a growing research on the potentials for “urban mining.”

The studies have also contributed to the increase in the understanding of the development of the urban flows in different cities, which often have changed dramatically during the twentieth century. Where city development has been characterized by increasing welfare, consumption, and mobility, the societal flows have grown dramatically and become more diversified, and spatial patterns have changed with the development and expansion of city regions. Most studied cities have, in recent decades, become more energy and material intensive, while other flows, such as water and wastewater, as well as different types of pollution show varying trends in different cities [27]. While Asian cities may still be characterized by increasing levels of traditional pollutants, such as SO₂ and particulates, European cities have generally experienced a dramatic trend break of most of such traditional air pollutants, while emissions of NO_x and CO₂ may remain stable [10]. In parts of East, South, and Central Asia, urbanization has dramatically increased water consumption and water pollution, which adds pressure on scarce water resources in the regions.

Urban Metabolism Studies and Planning

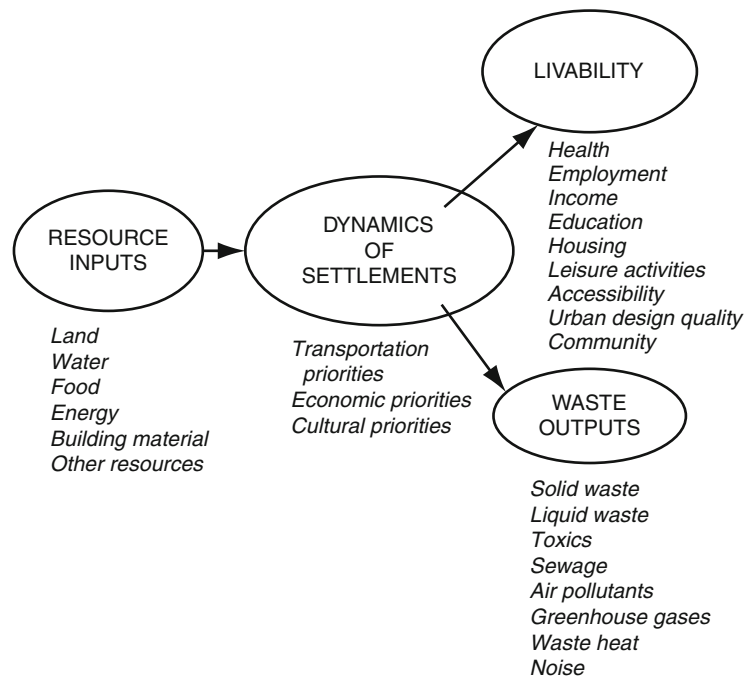
Urban metabolism is often described as “a holistic approach to urban planning” [5] and is fundamental for developing sustainable cities and communities. Such overviews of flows in different cities have, in many cases, been important for informing and stimulating strategic discussions concerning resource provision, resource efficiency, and pollution. Metabolism studies contributed importantly to the development of flow-based ecocycle strategies and related introduction and campaigns for recycling in the 1990s. There are examples of collaboration between authorities and researchers in some studies e.g., [28] and that studies

have been used for developing local sustainability indicators connected to energy or material consumption or ecological footprints. There are also examples of policy frameworks that have been developed for supporting sustainable urban policy development, e.g., the extended metabolism for human settlements from Australia that link the urban metabolism to both socio-economic dynamics and quality of life (Fig. 4). However, it is still rare that urban metabolism studies and perspectives have been fully integrated into local strategies or planning processes.

Kennedy et al. [27] suggest that particularly relevant practical applications of urban metabolism for urban planners and designers include sustainability reporting, urban greenhouse gas accounting, mathematical modeling for policy analysis, and urban design. Urban metabolism studies have often contributed to sustainability reporting and indicators. Indicators based on metabolism studies have many advantages: they are not only indicative of a city's sustainability but also are scientifically valid, relevant to urban planners and dwellers, comparable over time, and understandable. With the growing interest for setting local goals

for emissions of greenhouse gases and ecological footprints, metabolism approaches are needed for estimating the emissions in different cities. An increasing number of cities – from the planned city of Masdar in Abu Dhabi to big cities such as Copenhagen and Seattle – have, in the last few years, declared ambitions to become carbon neutral within the next few decades. Partly connected is a trend among “Sustainable Cities” to set targets in terms of decreasing their ecological footprints.

The ambitions of metabolism studies to contribute to regional assessment, modeling, and scenario construction have only seldom been realized, though there are exceptions, such as City Region 2020 [4] and Maastricht 2030 [5], assessment and scenario studies based on metabolism approaches. There are also examples of mathematical models of metal and nutrient flows, such as SIMBOX [63] and STAN [47, 64], that can be used for simulating future changes and identify solutions to perceived problems. Urban metabolism has furthermore been used for development of tools for designing and redesigning urban areas and their infrastructure [27, 65].



Natural Resource Flows and Sustainability in Urban Areas. Figure 4
Extended metabolism model of human settlements (From: [7])

Limitations and Challenges

The metabolism studies have traditionally limitations that can be connected to data availability, perspectives, and available resources and priority setting in individual projects. Observations and discussion about the difficulties of the studies in terms of data accessibility, missing information, and uncertainties are often neglected in presentations of various studies. Statistics do not adapt very well for following all types of flows and particularly not in a local area. Many types of statistical data and inventories are often only available at the national level, and these data are often used as basis for estimating local flows. Data on water, energy, and waste flows handled by local public organizations are more accessible, which explains the dominance of these in studies. With the exception of very few city areas, such as Hong Kong, which has been subject to repeated studies [66–68], cities are difficult to delimit, which creates problems to estimate inflows to the urban system as well as their international and local exchanges.

Integrating nature and socioeconomic aspects is still an important challenge in connection with ecosystem analysis. The analysis of urban metabolism has largely been dominated by quantification of material inflows and outflows, and the connections to social aspects have often been fairly weak. Only few studies have made efforts to analyze flows in their broader social, economic, and geographical setting, and studies have seldom been able to connect to decision-making or involve stakeholders. Despite increased interest in international aspects of material flows, only limited systematic efforts to evaluate ecological linkages of various cities in more detail than standardized ecological footprints have been made. It is also rare with several levels of scale in the analysis. Summarized flows contribute to an overview but are not adequate for making connections to decision points in society or say much about their qualitative usefulness to society [3].

In making connections to decision-making, planning, and strategy development, it is necessary with closer connections to decision points and institutional structures, and the analysis must become more flexible in scale and detail. A more systematic multilevel analysis and connections between different levels of scale

and governance are needed. In many cities, there is a lack of evaluation of local ecocity projects in relation to the whole city and its development. In a discussion paper on how cities can integrate global concerns, Xia Bai [69] presented some thoughts on how scale considerations should be taken into account more systematically when addressing city strategies, which can inspire urban metabolism studies to become more policy relevant.

Future Directions

Many innovative ideas regarding how a city could be planned or developed to minimize its resource consumption have been proposed in recent decades. There are numerous ecocity or sustainable city initiatives in different cities around the world. Cities such as Curitiba, Freiburg, London, Vancouver, Stockholm, Tianjin, and Melbourne can show inspiring examples of initiatives and positive developments. Public planning authorities have developed growing awareness of sustainability issues, and it is difficult to find any comprehensive urban plan that lacks sustainability goals. In local ecocity projects, flow-based initiatives are common in relation to waste, energy, and traffic. Despite this marked increase in awareness and ambitions, there is still a lack of powerful initiatives that decisively shift urban development in a sustainable direction. Even cities that have a high profile as sustainability forerunners, such as Copenhagen and Malmö [70] or Portland [71, 72], may show developments that are fairly inconsistent with declared sustainability ambitions, particularly in terms of road traffic and urban sprawl.

There is generally a need for more comprehensive analysis at the city and big city region level by revitalized urban metabolism efforts. It is, however, possible that the increasing interest among cities concerning climate change mitigation and ecological footprints will stimulate further analysis of the urban metabolism at the city level.

Urban metabolism and ecological footprints have become influential concepts. But it is a bit paradoxical that while urban ecology and sustainable urban development research and actions have grown dramatically, the analysis of urban systems at the city level or big city regional level has played only a marginal role. Efficiency

of energy, problems of traffic flows, and consumption are addressed in an urban setting, but the scale is most often far below the city or city region level (e.g., house, household, city district, infrastructural sector). There exists seldom a sufficiently good information basis for assessing the development of the city and its region and their external exchanges with the surrounding world and for developing comprehensive strategies at the city level.

There are many intriguing questions concerning the development of big city regions, e.g., connected to urban form and location policies that should be systematically analyzed from a metabolic perspective. From studies primarily in the UK and the Netherlands, there is evidence of positive effects in terms of provisions of ecoservices and minimized loss of natural environment and biodiversity from promotion of more compact urban development [73, 74]. However, there is still debate about the desirability and challenges of such compact city ambitions e.g., [75] and need for more empirical evidence [3].

In dynamic big city regions that are characterized by regional enlargement and polycentric developments, such issues are very complex and challenging but need to be analyzed further. In national planning systems all over the EU, the concept of polycentricity is implemented [76]. The European Spatial Development Perspective [77] strongly argued that polycentric urban systems are more equitable, efficient, and sustainable than monocentric or dispersed urban systems and should therefore be actively encouraged, but this hypothesis remains to be tested, and even if polycentric regions have certain potential advantages, these opportunities still need to be identified and used. Major research questions in both the large urban metabolism projects under EU-FP7 connect to these issues. They both focus on urban form, traffic, and land use, and their ambitions to connect to planning and institutional structures are higher and more elaborated than most past urban metabolism projects.

There are many challenges for future urban metabolism studies in relation to developing the overview and understanding of the urban flow systems and to become more policy relevant. Only about 30 cities have been so far analyzed, and there have not been much comparative analysis, generalization, and theoretical development. Kennedy et al. [27] suggested that

identification of various classes and comparative analysis of various factors, such as climate, age of city, and development history, should be further focused. Comparative studies are difficult due to very different preconditions in various cities, and comparisons between existing studies are problematic due to inconsistent basic data and lack of standardization of procedures. Even though there has been standardization effort in some study areas (e.g., metal metabolism studies), such challenges are likely to remain. Despite this, comparative studies and comparisons with other cities may still be valuable and almost necessary for developing perspectives on different resource flows and their efficiency in individual cities. Comparative studies can also form a base for looking closer into driving forces of the urban metabolism and effects of different developments or strategies. Besides improving the assessment of linkages of the urban metabolism in different scales, future studies should also aim to improve the overview of storage and accumulation processes in the urban environment. Analysis of urban natural flow systems is of central importance for a more comprehensive understanding of the resource use in the city and for assisting cities to contribute more forcefully to global sustainable development.

Bibliography

Primary Literature

1. UNFPA (2007) State of world population 2007. Unleashing the potential of urban growth. United Nations Population Fund, New York
2. Grimm NB, Faeth SH, Golubiewski NE, Redman CL, Wu J, Bai X, Briggs JM (2008) Global change and the ecology of cities. *Science* 319:756–760
3. Huang SL, Yeh CT, Chang LF (2010) The transition to an urbanizing world and the demand for natural resources. *Curr Opin Environ Sustain* 2:136–143
4. Ravetz J (2000) Integrated assessment for sustainability appraisal in cities and regions. *Environ Impact Assess Rev* 20:31–64
5. Rotmans J, van Asselt MBA, Vellinga P (2000) An integrated framework for sustainable cities. *Environ Impact Assess Rev* 20:265–276
6. Alberti M (1996) Measuring urban sustainability. *Environ Impact Assess Rev* 16(4–6):381–423
7. Newman PWG (1999) Sustainability and cities: extending the metabolism model. *Landsc Urban Plan* 44:219–226
8. Batty M (2008) The size, scale, and shape of cities. *Science* 319:769–771

9. Luckin B, Massard-Guilbaud G, Schott D (eds) (2005) *Environment and the City. Modern European cities and the management of their resources*. Ashgate, Aldershot
10. Douglas I, Hodgson R, Lawson N (2002) *Industry, environment and health through 200 years in Manchester*. *Ecol Econ* 41(2):235–255
11. Keiner M, Koll-Schretzenmayr M, Schmid WA (2005) *Managing urban futures: sustainability and urban growth in developing countries*. Ashgate Publishers, Aldershot
12. Ooi GL (2009) Challenges of sustainability for Asian urbanization. *Curr Opin Environ Sustain* 2009(1):187–191
13. Agudelo-Vera CM, Mels AR, Keesman KJ, Rijnaarts HH (2011) Resource management as a key factor for sustainable urban planning. *J Environ Manage* 92:2295–2303
14. Roseland M (2005) *Toward sustainable communities: resources for citizens and their governments*. New Society, Gabriola Island
15. Newman P, Jennings I (2008) *Cities as sustainable ecosystems: principles and practices*. Island, Washington DC
16. Girardet H (1999) *Creating sustainable cities*. Green Books/The Schumacher Society, Totnes
17. Register R (2006) *Ecocities: rebuilding cities in balance with nature*. New Society, Gabriola Island
18. McIntyre NE, Knowles-Yáñez K, Hope M (2008) Urban ecology as an interdisciplinary field: differences in the use of “Urban” between the social and natural sciences. In: Marzluff JM (ed) *Urban ecology. An international perspective on the interaction between humans and nature*. Springer, New York, pp 49–65
19. Pickett STA, Cadenasso ML, Grove JM, Boone CG, Groffman PM, Irwin E, Kaushal SS, Marshall V, McGrath BP, Nilon CH, Pouyat RV, Szlavecz K, Troy A, Warren P (2011) Urban ecological systems: Scientific foundations and a decade of progress. *J Environ Manage* 92:331–362
20. Hultman J (1993) Approaches and methods in urban ecology. *Geograf Ann Ser B, Hum Geogr* 75(1):41–49
21. McGranahan G, Marcotullio P, Bai X, Balk D, Braga T, Douglas I, Elmqvist T, Rees W, Satterthwaite D, Songsore J, Zlotnick H, Eades J, Ezcurra E, Whyte A (2005) *Urban systems. In: Millennium ecosystem assessment: Conditions and trends assessment*. Island, Washington
22. Duvinéaud P, Denayer-De Smet S (1977) L'Ecosystème Urbain. In: L'Ecosystème Urbain Bruxellois. In: Duvinéaud P, Kestemont P (eds) *Productivité en Belgique. Travaux de la Section Belge du Programme Biologique International*, Bruxelles, pp 581–597
23. Wolman A (1965) The metabolism of cities. *Sci Am* 213: 179–190
24. Kennedy CA, Cuddihy J, Engel Yan J (2007) The changing metabolism of cities. *J Ind Ecol* 11(2):43–59
25. Ayres RU, Simonis UK (eds) (1994) *Industrial metabolism*. United Nations University Press, Tokyo
26. Newman PWG (1999) Sustainability and cities: extending the metabolism model. *Landsc Urban Plan* 44:219–226
27. Kennedy C et al (2011) The study of urban metabolism and its applications to urban planning and design. *Environ Pollut* 159:1965–1973
28. Bergbäck B, Johansson K, Mohlander U (2001) Urban metal flows—a case study of Stockholm, review and conclusions. *Water Air Soil Pollut Focus* 1(3/4):3–24
29. Barles S (2007) *Feeding the city: food consumption and flow of nitrogen, Paris, 1801–1914*. *Sci Total Environ* 375:48–58
30. Browne D, O'Regan B, Moles R (2009) Assessment of total urban metabolism and metabolic inefficiency in an Irish city-region. *Waste Manag* 29(10):2765–2771
31. Zucchetto J (1975) Energy, economic theory and mathematical models for combining the systems of man and nature. Case study, the urban region of Miami. *Ecol Model* 1:241–268
32. Huang SL, Hsu WL (2003) Materials flow analysis and energy evaluation of Taipei's urban construction. *Landsc Urban Plan* 63(2):61–74
33. Zhang Y, Yang Z, Yu X (2009) Evaluation of urban metabolism based on energy synthesis: a case study for Beijing. *Ecol Model* 220(13–14):1690–1696
34. Ascione M, Campanella L, Cherubini F, Ulgiati S (2009) Environmental driving forces of urban growth and development. An emergy-based assessment of the city of Rome, Italy. *Landsc Urban Plan* 93:238–249
35. Wackernagel M, Kitzes J, Moran D, Goldfinger S, Thomas M (2006) The ecological footprint of cities and regions: comparing resource availability with resource demand. *Environ Urban* 18(1):103–112
36. Wackernagel M, Rees W (1995) *Our ecological footprint: reducing human impact on the earth*. New Society, Gabriola Island
37. Wackernagel M (1998) The ecological footprint of Santiago de Chile. *Local Environ* 3(1):7–25
38. Folke C, Jansson Å, Larsson J, Costanza R (1997) Ecosystem appropriation by cities. *Ambio* 26:167–172
39. Barrett J, Vallack H, Jones A, Haq G (2002) *A material flow analysis and ecological footprint of York*. Technical report. Stockholm Environment Institute, Stockholm
40. Collins A, Flynn A, Weidmann T, Barrett J (2006) The environmental impacts of consumption at a subnational level: the ecological footprint of Cardiff. *J Ind Ecol* 10(3):9–24
41. Chartered Institute of Wastes Management (2002) *A resource flow and ecological footprint analysis of greater London*. Best Foot Forward, London
42. Eaton RL, Hammond GP, Laurie J (2007) *Footprints on the landscape: an environmental appraisal of urban and rural living in the developed world*. *Landsc Urban Plan* 83:13–28
43. Niza S, Rosado L, Ferrão P (2009) Urban metabolism: methodological advances in urban material flow accounting based on the Lisbon case. *J Ind Ecol* 13(3):384–405
44. Schulz NB (2007) The direct material inputs into Singapore's development. *J Ind Ecol* 11(2):117–131
45. Sörme L, Bergbäck B, Lohm U (2001) Century perspective of heavy metal use in urban areas. A case study in Stockholm. *J Water Air Soil Pollut Focus* 1(3–4):197–211

46. Svidén J, Jonsson A (2001) Urban metabolism of mercury turnover, emissions and stock in Stockholm 1795–1995. *J Water Air Soil Pollut Focus* 1(3–4):79–196
47. Brunner PH, Rechberger H (2004) Practical handbook of material flow analysis. CRC Press, Boca Raton
48. Obernosterer R, Brunner PH (2001) Urban metal management the example of lead. *J Water Air Soil Pollut Focus* 1(3–4): 241–253
49. Baker LA, Hope D, Xu Y, Edmonds J, Lauver L (2001) Nitrogen balance for the Central Arizona Phoenix (CAP) ecosystem. *Ecosystems* 4:582–602
50. Færgé J, Magid J, Penning de Vries FWT (2001) Urban nutrient balance for Bangkok. *Ecol Model* 139:63–74
51. Forkes J (2007) Nitrogen balance for the urban food metabolism of Toronto. *Can Resour Conserv Recycl* 52:74–94
52. Neset TSS, Lohm U (2005) Spatial imprint of food consumption. A historical analysis for Sweden, 1870–2000. *Hum Ecol* 33(4):565–580
53. Hermanowicz SW, Asano T (1999) Abel Wolman's the metabolism of cities' revisited: a case for water recycling. *Water Sci Technol* 40(4):29–36
54. Gandy M (2004) Rethinking urban metabolism: water, space and the modern city. *City* 8(3):363–379
55. Thériault J, Laroche AM (2009) Evaluation of the urban hydrologic metabolism of the Greater Moncton region, New Brunswick. *Can Water Resour J* 34(3):255–268
56. Sahely HR, Kennedy CA (2007) Integrated systems flow model for quantifying environmental and economic sustainability indicators: case study of the City of Toronto urban water system. *ASCE J Water Resour Plann Manage* 133(6):550–559
57. Baker LA (ed) (2009) The water environment of cities. Springer, New York
58. Kennedy C, Steinberger J, Gasson B, Hillman T, Havránek M, Hansen Y, Pataki D, Phdungsilp A, Ramaswami A, Villalba Mendez G (2009) Greenhouse gas emissions from global cities. *Environ Sci Technol* 43:7297–7302
59. Fischer-Kowalski M, Krausmann F, Smetschka B (2004) Modelling scenarios of transport across history from a socio-metabolic perspective. *Rev – Fernand Braudel Center* 27:307–342
60. Bettencourt L, Lobo J, Helbing D, Kuhnert C, West GB (2007) Growth innovation scaling and the pace of life in cities. *Proc Natl Acad Sci USA* 104:7301–7306
61. SUME (Sustainable Urban Metabolism for Europe) FP7 Collaborative Research Project. <http://www.sume.at>. Accessed 1 Jul 2011
62. BRIDGE programme (sustainable urban planning Decision support accountinG for urban metabolism). <http://www.bridge-fp7.eu>. Accessed 1 Jul 2011
63. Baccini P, Bader HP (1996) Regionaler Stoffhaushalt. Spektrum Akad. Verlag, Heidelberg
64. Cencic O, Rechberger H (2008) Material flow analysis with software STAN. *J Environ Eng Manage* 18(1):3–7
65. Oswald F, Baccini P (2003) *Netzstadt: designing the urban*. Birkhäuser, Basel
66. Newcombe K, Kalma J, Aston A (1978) The metabolism of a city: the case of Hong Kong. *Ambio* 7:3–15
67. Boyden S, Millar S, Newcombe K, O'Neill B (1981) The ecology of a city and its people: The case of Hong Kong. Australian National University Press, Canberra
68. Warren-Rhodes K, Koenig A (2001) Escalating trends in the urban metabolism of Hong Kong: 1971–1997. *Ambio* 30(7):429–438
69. Bai X (2007) Integrating global environmental concerns into urban management – the scale and readiness arguments. *J Ind Ecol* 2(11):15–29
70. Anderberg S, Clark E (2011) The green and sustainable Øresund region: Eco-branding Copenhagen and Malmö. In: Vojnovic I (ed) *Sustainability: a global urban context*. Michigan State University Press, Lansing (in press)
71. Redkin A Portland again tops a sustainable cities list, dot earth, the opinion pages, New York Times. <http://dotearth.blogs.nytimes.com>. Accessed 22 Jun 2011
72. Sustainable Business Oregon Portland's traffic is 16th worst. <http://www.sustainablebusinessoregon.com>. Accessed 22 Jun 2011
73. Tratalos J, Fuller R, Warren P, Davies R, Gaston K (2007) Urban form, biodiversity potential and ecosystem services. *Landsc Urban Plan* 83:308–317
74. Geurs KT, van Wee B (2006) Ex-post evaluation of thirty years of compact urban development in the Netherlands. *Urban Stud* 43(1):139–160
75. Jenks M, Burton E, Williams K (eds) (2000) *The compact city: a sustainable urban form?* Oxford Brookes University, Oxford
76. Halbert L, Pain K, Thierstein A (2006) European polycentricity and emerging Mega-City-Regions – “one size fits all” policy? *Built Environ* 32(2):194–218
77. European Commission (1999) *European spatial development perspective. Towards balanced and sustainable development of the territory of the European Union*

Books and Reviews

- Anderberg S, Prieler S, Olendrzynski K, de Bruyn S (2000) *Old sins – industrial metabolism, heavy metal pollution and environmental transition in Central Europe*. UN University Press, Tokyo
- Baccini P, Brunner PH (1991) *Metabolism of the anthroposphere*. Springer, Berlin
- Bernhardt C (ed) (2004) *Environmental problems in European cities in the nineteenth and twentieth centuries*. Waxmann, Münster/New York
- Odum HT (1996) *Environmental accounting: emergy and environmental policy making*. Wiley, New York
- Tarr JA (2002) The metabolism of the industrial city: the case of Pittsburgh. *J Urban Hist* 28(5):511–545

Natural Ventilation in Built Environment

TONG YANG¹, DEREK J. CLEMENTS-CROOME²

¹Department of Civil and Building Engineering,
Loughborough University, Leicestershire, UK

²School of Construction Management and
Engineering, University of Reading, Reading, UK

Article Outline

Glossary

Definition of the Subject

Introduction

Vernacular Architecture

Natural Ventilation Principles

Natural Ventilation Design Requirements

Design Guidelines

Selection of Ventilation Strategies

Case Studies

Future Directions

Acknowledgments

Bibliography

Glossary

Air changes per hour (ACH) The volumetric flow rate of supply air, divided by the volume of the ventilated space.

Advanced natural ventilation system (ANV) Integration of basic natural ventilation strategies such as cross ventilation and stack effect with smart controls.

BEMS Building energy management system.

BREEAM Building research establishment environmental assessment method – UK origin.

Exfiltration/infiltration Air flow through unintended leakages out/into buildings.

Hybrid ventilation Combined natural and mechanical ventilation (also called mixed-mode ventilation).

Indoor air quality (IAQ) Indoor Air Quality – broadly defined by the purity of the air but often CO₂ is used as an indicator.

Mixed-mode ventilation See hybrid ventilation.

Natural ventilation Use of natural forces, i.e., pressure differences generated by wind or air temperature, to

introduce and distribute outdoor air into or out of a buildings.

Night cooling The use of night air to cool the building using wind towers or a fan to circulate the air.

PAQ Perceived air quality.

Thermal comfort The state of mind that expresses satisfaction with the surrounding thermal environment.

Ventilation Provides fresh air into a building to ensure good air quality for occupant health and well-being.

Ventilation effectiveness The ability of a ventilation system to exchange the air in the room and also the ability to remove airborne contaminants.

Ventilation flow rate The amount of air per unit time into the ventilated space (liter per second or l/s, cubic meters per hour or m³/h).

Well-being Healthy mind and body.

Definition of the Subject

Natural ventilation uses the natural forces of wind and buoyancy to introduce fresh air and distribute it effectively in buildings for the benefit of the occupants. Fresh air is required to achieve a healthy, fresh, and comfortable indoor environment for people to work and live in. Natural ventilation can ensure or support the supply of adequate breathing air, adequate ventilation of contaminants, adequate thermal conditioning and moisture dissipation, and contribute to well-being through a connection to the dynamics of nature. For natural ventilation to be effective, there has to be a close relationship between the architecture and the air circulation system. This includes the relationship between the built form, the site environment in a particular location, and the layout within the building.

The Natural History Museum in London, designed by Alfred Waterhouse in the Victorian age, is an excellent example of design for natural ventilation. The architect designed the built form to encourage the flow of air through each space in the building by the use of two ventilation towers at the back of the building to induce air flow through stack ventilation [1].

Buildings should be designed to take full advantage of the prevailing natural forces such as wind, outdoor temperature, sunlight, incorporating building elements such as towers, atria, and thermal mass to ventilate and cool occupied spaces. In many climates there is

a growing proportion of naturally ventilated buildings using natural features and forces to reduce a building's *environmental or carbon footprint*.

Introduction

The reasons for ventilating a space with air are as follows:

1. Ventilation air provides oxygen that is needed for human life processes; it takes about 4 s for inhaled air to pass through the respiratory system and transfer oxygen to the blood and then to the brain; poor-quality air deficient in oxygen with consequent high CO₂ levels impedes clear thinking and concentration.
2. Ventilation air dilutes; the contaminants may be CO₂ from respiration, odors secreted through the human skin, cigarette smoke, or emissions from other process such as dust, allergens, aerosols, toxic gases, and particulates in general.
3. Ventilation promotes and directs air movement in the space, removing excessive heat and/or moisture essential for comfort and well-being.

Traditional vernacular architecture has taught us the best of sustainable architecture and ecologically sensitive adaptation, using passive features ranging from building orientation and form, appropriately sized and oriented openings linked with vertical forms, the benefits of local materials and mass for night cooling, and the relationship of buildings in context to ensure effective air flows.

Vernacular Architecture

Vernacular architecture blends buildings into their specific settings, so that there is a natural harmony between the climate, architecture, and people. Vernacular architecture learned from the environmental variations of place relating to local variations in temperature, humidity, sun, wind, rain, earthquakes, and storms. In climates where the diurnal range may be 17°C, vernacular buildings allow a variation in indoor temperature of only some 4°C through time-lag and night cooling. In climates where humidity may be 90%, vernacular buildings support human comfort by allowing air to flow over the many thermo-receptors on the human body. Vernacular architecture also

adapted to ensure indoor air quality through natural ventilation through the careful design and placement of indoor pollutant generators from stoves to commodes. Four vernacular solutions are further described: wind towers and courtyards, termite mounds, and igloos, each integrating the conditioning power of natural ventilation in unique responses to local climate.

Wind Towers

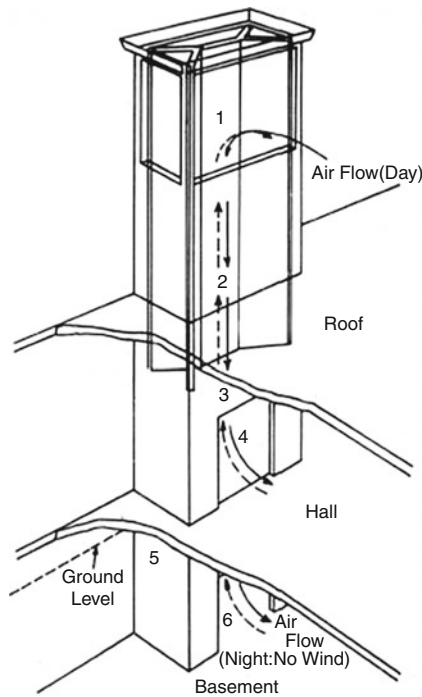
The wind towers called *bagdirs* are a distinctive and ancient feature of Islamic architecture. It has been used for centuries to create natural ventilation in buildings. Examples of wind towers (Fig. 1) can be found throughout the Middle East, Pakistan, and Afghanistan and now are sometimes incorporated into Western architecture.

Wind flowing around a building causes separation of flow which creates a positive pressure on the windward side and a negative pressure on the leeward side of the building. Due to its height the wind tower enhances the positive pressure on the windward side; it is then directed through the tower into the building. Airflow follows the pressure gradients within the structure and exits through purposely designed openings and as well as through the leeward side of the tower. The size and location of openings (e.g., windows, doors, etc.) and distribution of internal party walls have a great impact on encouraging cross flow and mixing of the indoor air.

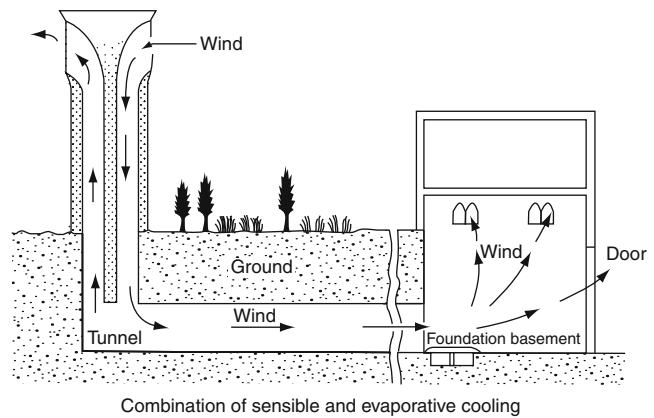
The principal factor is the buoyancy which depends on the temperature difference and the height. During the day the sun heats up the structure warming the internal air which then rises through the wind tower, as illustrated in Fig. 1. At night the cool night air lowers the temperature of the structure and the internal air and the heavier air then flows downward cooling the internal spaces after the heat of the day. Figure 2 shows how wind towers can also provide natural cooling for underground water cisterns.

Courtyards

Courtyards are one of the oldest plan forms for dwellings going back thousands of years and appearing as a distinctive form in many regions in the world. Examples exist in Latin America, China, the Middle East, Mediterranean, and in Europe. Preserving



Natural Ventilation in Built Environment. Figure 1
Bagdir in Dubai, in United Arab Emirates [2]



Natural Ventilation in Built Environment. Figure 2
Wind towers in Yazd, Iran to ventilate houses, are also constructed to cool underground cisterns (water reservoir) [3]

the basic typology of the courtyard, local climate and culture has created a unique style for each region.

The Courtyard House (Siheyuan) is a typical form in ancient Chinese architecture, especially in northern China. It offers space, comfort, quiet, and privacy.

A Siheyuan consists of a rectangle with a row of houses bordering each side around a courtyard, normally with a southern orientation and having the only gate usually situated in the southeast side. Walls protect the houses from the harsh winter winds and the spring dust storms

that frequently occur in Northern China from the Gobi desert in Mongolia. The house's deep eaves allow the winter sun's warmth to be directed into the rooms, while they also provide cooling shade and protection from the summer rains. Their design reflects the traditions of China, following the rules of Feng Shui and Confucian tenets of order and hierarchy.

All the rooms around the courtyard have doors and large windows facing onto the yard and small windows high up on the back wall facing out onto the street. Ridged roof tops provide shade in the summer and retain warmth in the winter. The verandah divides the courtyard into several big and small spaces that are closely connected, providing a common place for people to enjoy whatever the weather. The courtyard is an open-air living room and garden with plants, rocks, and flowers, for family members to chat and gather.

In cold northern China, courtyards are built broad and large to increase the exposure to sunlight, and there are more open areas inside the courtyard walls for daylight, fresh air, rainwater for plants and gardens to be harnessed. In hot southern China, the courtyard houses (Fig. 3) are built with multiple stories to encourage cross ventilation flow incorporating natural

cooling effects. The orientation of houses is not strictly north–south aligned, but follows the local topology of hills and easy access to water sources.

Lessons from Nature: Termitaries

Termites are an outstanding example in the animal kingdom of ingenious animal architects in the sense of master builders. Over 2,000 species live in tropical and subtropical regions and have shown us by analogy the art of designing for living in a variety of dwelling styles with natural ventilation.

Termites build their nest so as to achieve automatic ventilation to regulate the internal temperature, as well as constantly managing control of gas exchange and moisture level. They do not keep a set temperature, but allow a gradual change between the seasons determined by the external environmental temperature.

In Australia, compass termites build large-sized mounds in the form of huge, flat chisel-shaped blades, with their long axis pointing north–south. This arrangement exposes the minimum possible area to the midday sun but allows the mounds to catch the rays of the early morning and late evening sun, when the termites need warmth, especially in the cold season; peak temperatures can be lowered by about 7°C with N–S orientation and thus maintain a preferred temperature of 30–32°C [4].

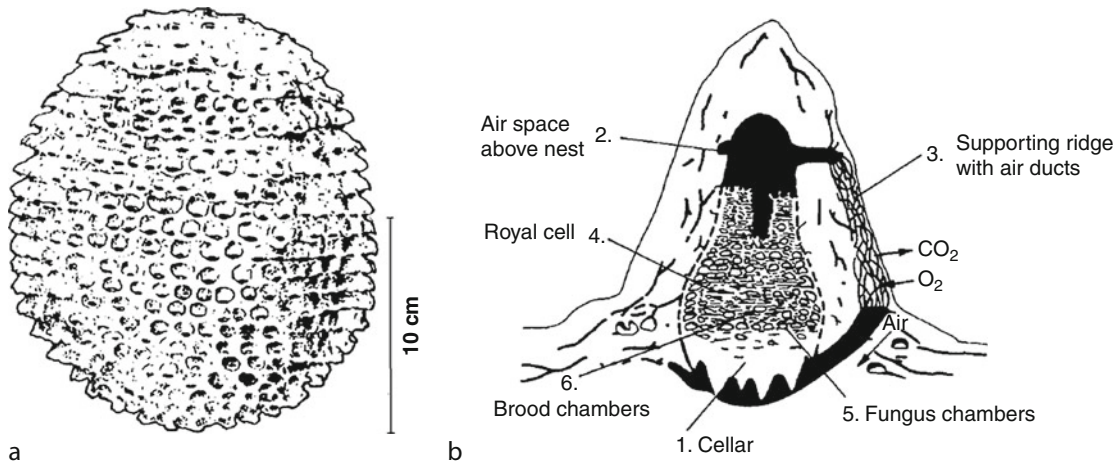
There are two main types of termite mounds: (1) The *open ventilation* mounds which let air flow into or out through chimneys or holes built into the mounds; (2) The completely *enclosed mounds* in which gases are exchanged through the porous thin-walled tunnels.

The nest of a termite species *Apicotermes gurgulifex* is shown in Fig. 4a. It is embedded in the soil but clothed by a mantle of air; the nest is constructed from the excrement of the termites so is well insulated. Its outer wall has a pattern of raised, ring-shaped configurations which surround an array of precisely spaced and shaped ventilation slits. These slits link the external and the internal spaces.

The termitary of the fungus cultivating termites, *Macrotermes bellicosus* in Fig. 4b, may reach a height of 3 or 4 m and contain more than two million termites. The fungus chambers are built into complex sponge-like structures with numerous supporting

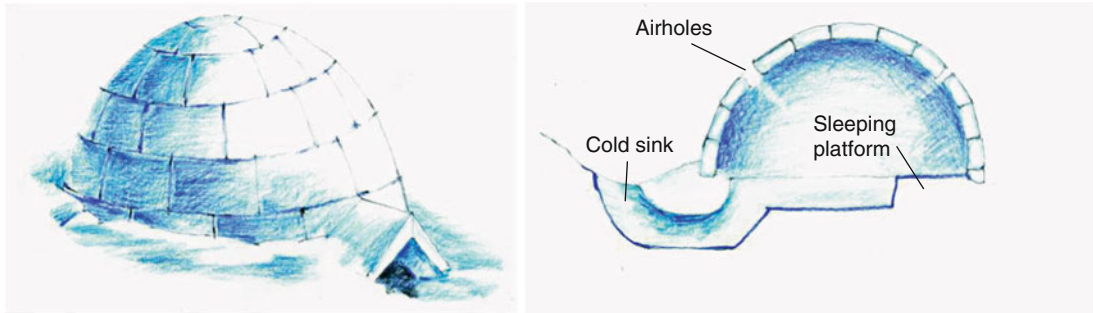


Natural Ventilation in Built Environment. Figure 3
A typical courtyard house in southern China



Natural Ventilation in Built Environment. Figure 4

Ventilation of termite molds (a) Nest of a termite species *Apicotermes gurgulifex* [4]; (b) Longitudinal section through the nest of *Macrotermes bellicosus* from Ivory Coast showing the air being circulated by buoyancy [4]



Natural Ventilation in Built Environment. Figure 5

Illustration of an igloo

ridges with air ducts. The air in the fungus chamber is heated by fermentation processes and the metabolic heat generated by the termites. The hot air rises and enters the duct systems in the ridges, the walls of which are porous allowing carbon dioxide to escape from and oxygen to enter the dwelling. The cooler air flows down to the cellar and replaces the rising warmer air.

Igloos

Inuit people build igloos as shelters from the extreme weather conditions in the Arctic. The igloo (Fig. 5) has an excellent thermal performance without mechanical

equipment. The hemispherical shape of the igloo provides the maximum resistance to winter gales from all directions, which at the same time exposing the minimum surface area to heat loss. The dome uses packed snow blocks, some 500 mm thick, 1,000 mm long, and 150 mm wide, which are laid in a continuous in sloping pile. Effectively, the shape encloses the largest volume with the least material, so it can be heated by a blubber lamp. Coated by a glaze of ice on the interior surface, the finished dome is made stronger and windproof. The interior surface is also draped with animal skins and furs to prevent radiant and convective heat loss between the cold floor and the walls. Measurements have shown that with no heat source apart from the

small blubber lamp, internal air temperatures are held at levels of -6 to 4°C with external temperatures of -24 and -40°C [5].

General

Traditional building technologies have evolved and been adapted over time by people and animals in all climates to meet specific needs, accommodating the values, economies, and the culture inspired ways of life. However versatile they are all reflect the basic principles described in the next section.

Natural Ventilation Principles

Natural forces to drive ventilation can be wind pressures or pressure generated by the density difference between indoor and outdoor air.

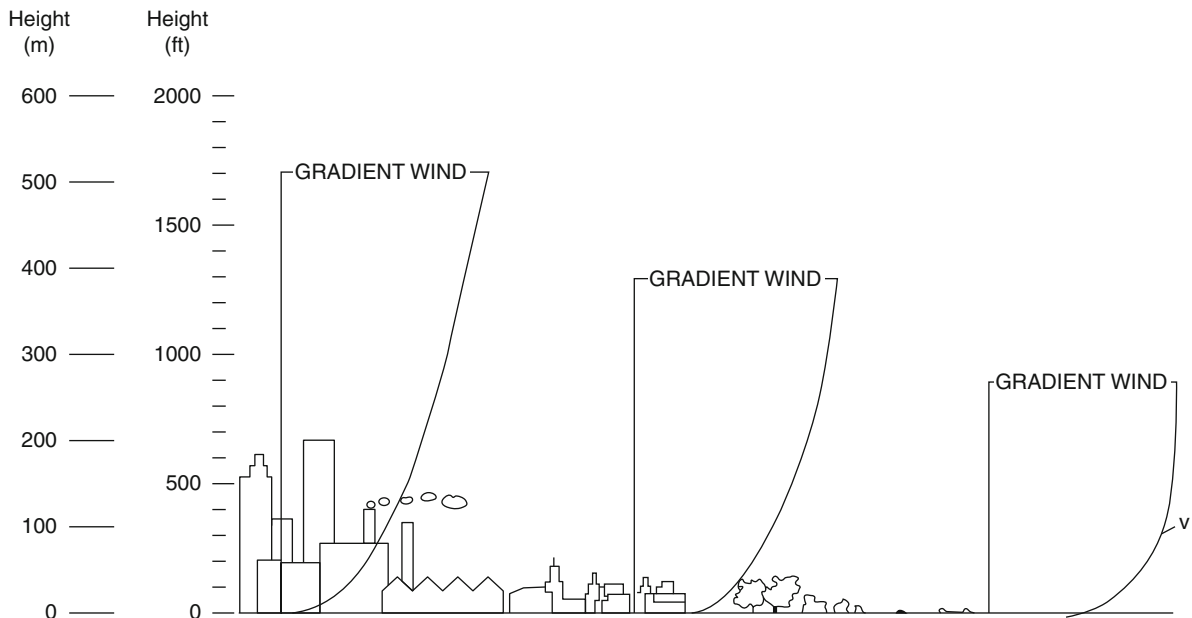
Wind-Driven Ventilation

Wind is caused by pressure differences in the atmosphere. The general flow of wind close to the Earth's surface is subjected to boundary layer effects, so called the atmospheric boundary layer, in which wind speed is

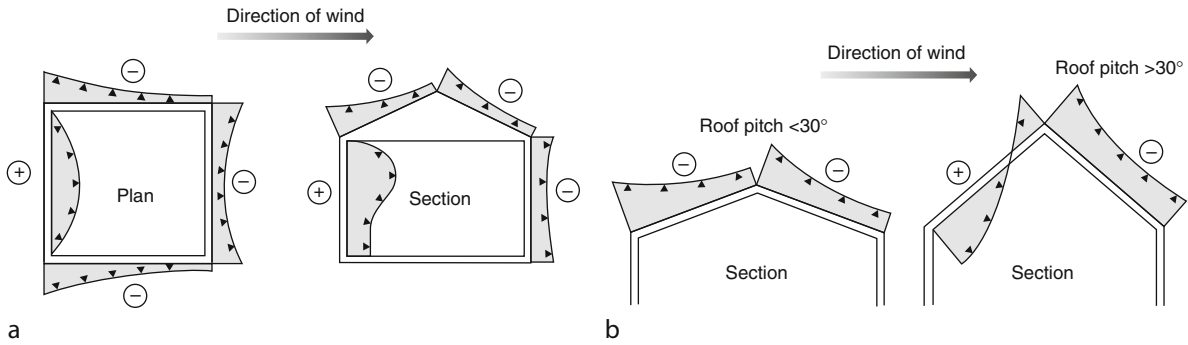
influenced by surface friction of the ground. The variation of wind speed in height on different terrains is illustrated in Fig. 6. Wind speed correction coefficients for different terrain conditions in the UK are listed in BS 5925 [6].

When the path of the wind is checked by obstacles, such as trees and buildings, then an energy conversion takes place. Velocity pressure is converted to static pressure, so that on the windward side an overpressure is produced (about 0.5–0.8 times the wind velocity), whereas on the leeward side an under-pressure results (about 0.3–0.4 times the wind velocity). The pressure distribution on the roof varies according to pitch. Figure 7 shows areas of positive and negative pressures generated by wind normal to building front: wind-driven flow through inlets on positive pressure faces and outlets on negative pressure faces [8]. The pressure differentials arising across a building cause infiltration of air through window cracks and other openings.

Relative to the static pressure of the free wind, the pressure on any point on the surface of a building façade p_w can be approximated by the equation:



Natural Ventilation in Built Environment. Figure 6
Wind speed variation with height and terrain conditions [7]



Natural Ventilation in Built Environment. Figure 7

Wind pressure distribution on buildings (a) Wind pressure on building; (b) Wind pressure on roof [8]

$$p_w = 0.5\rho C_p v_z^2 (P_a) \quad (1)$$

where

C_p = Wind pressure coefficient (dimensionless)

v_z = Local wind speed at a specific reference height z (m/s)

ρ = Air density (kg/m³)

In order to calculate the ventilation rate due to the wind a knowledge of the wind speed and direction is necessary besides information about the nature of wind-stream patterns developed around the building. Summary of façade-averaged wind pressure coefficient data for simple rectangular-plan low-rise buildings in varying degrees of shelter and wind directions is given in the *AIVC Application Guide: A Guide to Energy Efficient Ventilation* [9].

For buildings with a simple layout, the natural ventilation airflow rate by wind effect can be determined as in the following case, more examples are given in *CIBSE Applications Manual AM 10* [10]:

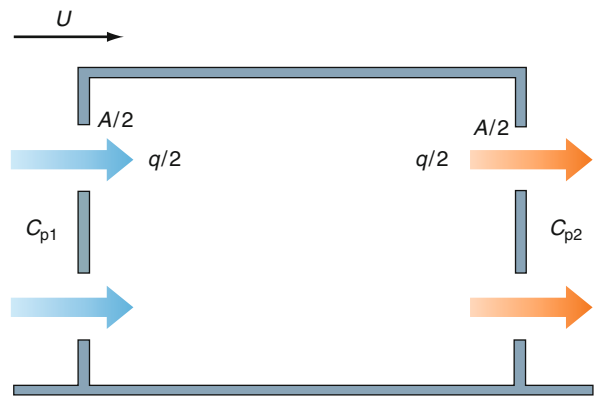
(a) Wind effect alone for a zone with multi cross flow openings (Fig. 8):

$$q = A \cdot C_d \cdot U \cdot \left[\frac{\Delta C_p}{2} \right]^{0.5} \quad (2)$$

where

U is the wind speed measured at the same height as the building (m/s)

A is the total ventilation area (m²) – assuming the four openings are identical



Natural Ventilation in Built Environment. Figure 8

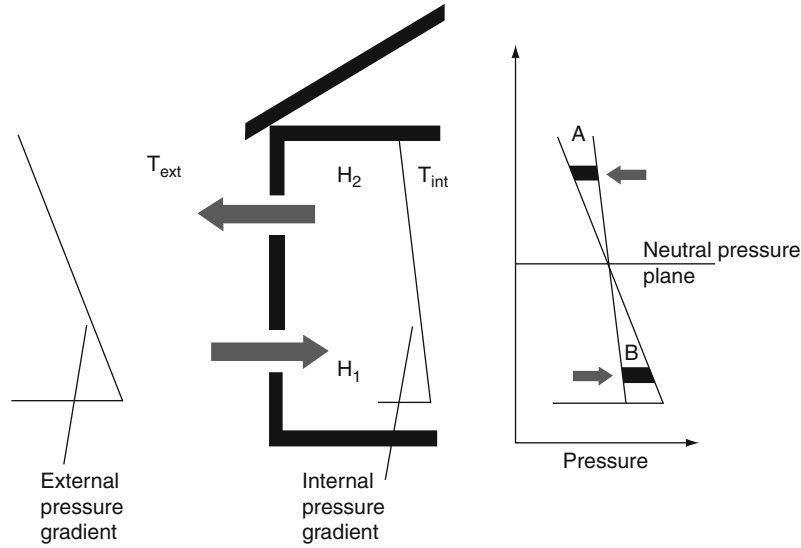
Cross ventilation driven by wind effect alone

C_d is the discharge coefficient (typical value ~ 0.6)

ΔC_p is the difference between wind pressure coefficient (C_{p1} and C_{p2})

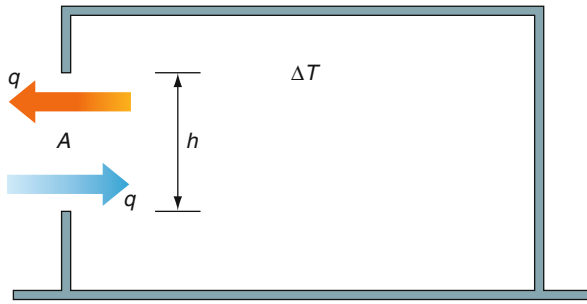
Buoyancy-Driven (Stack) Ventilation

Warm air in a room tends to rise because of its low density. It is replaced by cooler, denser air from outside. There is a neutral pressure plane where the pressure difference is zero as shown in Fig. 9. Since the pressure at the outlet or inlet can be affected by the wind, the extent to which the stack effect operates is governed partly by the wind pressure and partly by the design of the openings and the internal layout.



Natural Ventilation in Built Environment. Figure 9

Stack pressure distribution between two vertically placed openings ([9] p. 214)



Natural Ventilation in Built Environment. Figure 10

Single-sided ventilation, single-opening, driven by buoyancy alone

For buildings with a simple layout, the natural ventilation airflow rate by buoyancy effect can be determined as in the following case [10]:

(a) Buoyancy effect alone for a single-opening zone (Fig. 10):

$$q = \frac{A \cdot C_d}{\left[\frac{(T_i + 273)}{\Delta T \cdot g \cdot h} \right]^{0.5}} \quad (3)$$

where

q is the ventilation flow rate (m^3/s)

T_i is the internal temperature ($^{\circ}\text{C}$)

ΔT is the difference between the internal and external air temperature (K)

A is the opening area (m^2)

h is the opening height (m)

C_d is the discharge coefficient (~ 0.25 for single-opening)

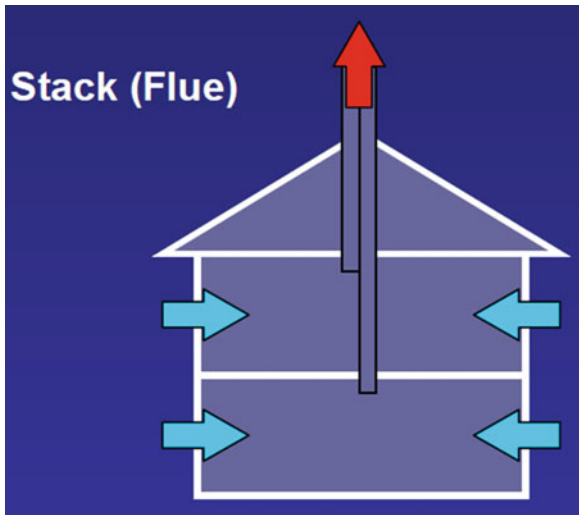
g is acceleration due to gravity (m/s^2)

In spaces with high ceilings and where cross ventilation is not feasible, stack ventilation works best (Fig. 11).

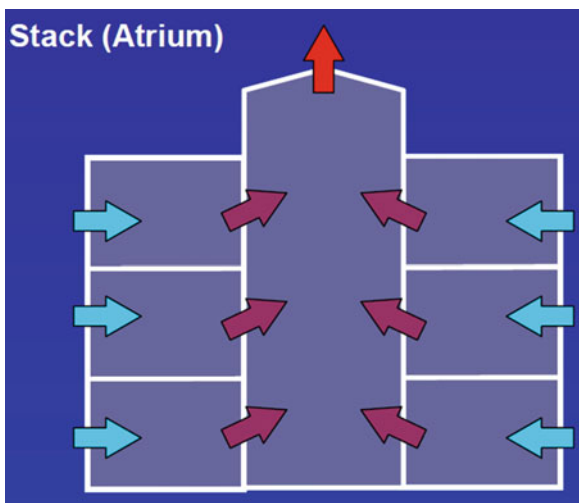
An atrium is a variant of stack ventilation principle (Fig. 12). It draws air from both sides of the building toward a central extract point, doubled the natural ventilation effective width in the building.

A well-designed double-skin façade provides buildings with additional protective layer from outdoor environment, improves thermal and visual comfort for occupants and reduce intrusive noise. It can create the circulation and integration of the internal and external space of building, providing views to outside and so achieve the harmonious integration of people and nature.

The Gherkin building (see Fig. 13), 30 St Mary Axe in the city of London, is naturally ventilated in most time of the year, its exterior cladding consists of triangular and diamond shaped glass panels. The glazing



Natural Ventilation in Built Environment. Figure 11
Stack ventilation from vertical chimneys [11]



Natural Ventilation in Built Environment. Figure 12
Stack ventilation from Atrium [11]

system contains a double-glazed outer layer and a single-glazed inner screen with solar-control blinds in the central ventilated cavity. Fresh air is drawn up through spiraling light wells which consists of openable double-glazed panels, also effectively reduce the need for additional heating and cooling. The circular tapering shape of the building and the light wells maximized the amount of natural light throughout the building and provide views out across the City from deep inside.

When natural forces cannot provide the required indoor environment conditions, mechanical systems/ components, – e.g., fans for increasing ventilation rate, and/or heat exchanger for cooling (or heating) at peak summer (or winter) times – can be used to enhance the natural ventilation through purposely installed openings in the building envelope. *Hybrid ventilation* (mixed-mode) system with pressure sensors and motor-driven dampers are used to give control. Careful considerations in the design and operation as well as end-user education are needed to deliver effective environmental control with ventilation.

Natural Ventilation Design Requirements

Fresh Air

Indoor CO₂ source is primarily generated through human metabolism. Human respiration requires a fresh air rate of 0.1–0.9 l/s per person depending on the activities (metabolic rate) and clothing insulation [14].

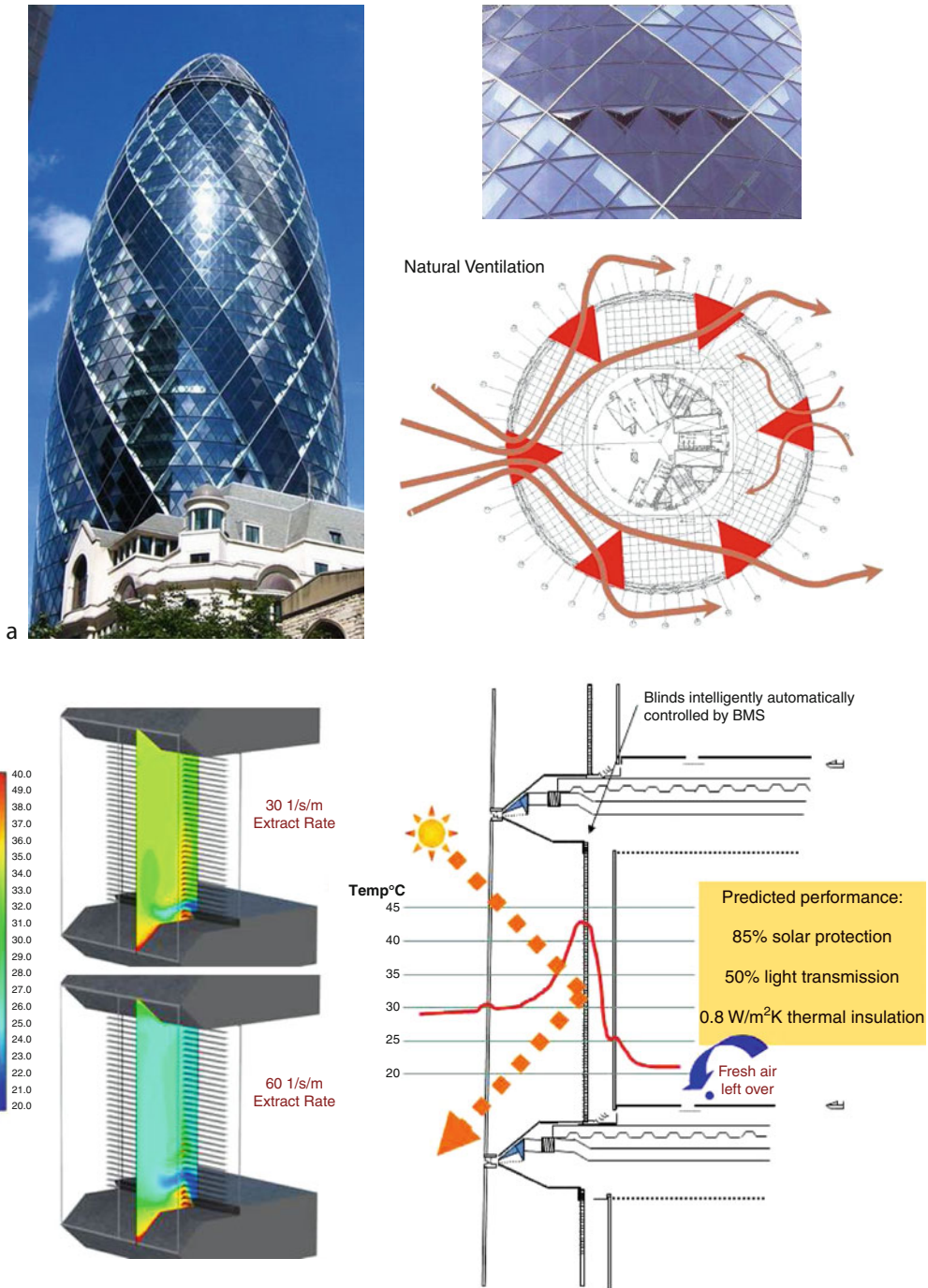
Seppänen et al. [15] and Wargocki et al. [16] have made a comprehensive review of over 20 studies with over 30,000 persons and found that ventilation rates below 10 l/s per person results in lower air quality and increase health problems.

Ventilation rates for acceptable indoor air quality are currently assessed by using the ASHRAE Standard 62.1 [17]. In this Standard there are two procedures for estimating the amount of fresh air required. The first is referred to as the ventilation rate and is a prescriptive approach stating that, for office buildings, there is a requirement of 10 l/s per person of fresh air (nonsmoking). A comparison of Standards for the indoor environment is given in the ASHRAE Handbook on Fundamentals [18].

Air Movement in Rooms

It has been well established that air movement is one of the important factors that influence people's perception of thermal comfort [19].

Air movement is a combination of a momentum-induced air flow and buoyancy-induced air flow [20]. A jet air flow is caused by a momentum source, which can either be a fan or the pressure difference across an opening caused by the wind or temperature difference.



Natural Ventilation in Built Environment. Figure 13

Gherkin London and ventilation through double-skin facade **(a)** Gherkin façade [12] and natural ventilation concept [13]; **(b)** Gherkin double-skin facade design [13]

The other main source of air motion is buoyancy-driven air flow caused by density differences. This type of motion is also called natural convection.

The research work of Linke [21], Mullejšans [22], and van Gunst [23] has given clear indications about the air patterns produced by air streams at various velocities and temperatures, when directed through different types of outlet, and also their interaction with the natural convection currents in the space. Optimum air and temperature distribution as well as a satisfactory sound level should be provided from the air streams outlet.

- Air movement should vary in space and time without giving drafts, especially, some parts of the body (i.e., ankles, back of the neck) are more susceptible to drafts.
- Temperatures should vary within a vertical gradient limit, higher level of warmth being preferable at below knee level rather than at head level.
- For freshness higher air velocity are required at higher temperatures, an air velocity change of 0.15 m/s being equivalent to a change of about 1°C in temperature. Air at a lower temperature and relative humidity of 40–60% (i.e., air with a lower enthalpy is perceived as fresher than air with a higher enthalpy [24, 25]).
- Above the head the convection air velocities can be 0.25 m/s or higher depending on the occupancy density and also the amount of artificial lighting.
- Air movement helps to dispel a sense of stuffiness.

Indoor Air Quality and Natural Ventilation

Body Odors A human being's sense of smell permits very low concentrations of odors. The sensitivity varies between individuals. In a typical indoors around 500 out of 6,000 compounds are human bioeffluents [26].

Poor-quality air is usually referred to as being stuffy, stale, close, heavy, or lacking in freshness. Inhaled air comes into contact with the nasal passages and then the respiratory tissue; in each case the motion of the olfactory hairs, and of the cilia on the respiratory membrane, is affected by temperature, humidity, dirt, odors, and also ions [27].

Environmental Tobacco Smoke Environmental tobacco smoke contains more than 4,000 chemicals

and at least 50 of these chemicals are known cancer-causing substances [28]. It generates about 2 ppm CO, leads to irritation and discomfort among 20% of those exposed, also suspected increase in the risk of lung cancer [29]. Based on a study on the rates of hospital admissions for heart attacks before and after the smoking ban was introduced in England on July 01, 2007, there was a clear association between the smoking ban and a 2.4% reduction (or 1,200 fewer admissions due to a heart attack) in the 12 months following the ban [30].

Metabolic Carbon Dioxide An average sedentary adult (Metabolic Rate $M = 70 \text{ W/m}^2$ and body area $A = 1.8 \text{ m}^2$) produces approximately 0.0051 l/s (18 l/h) of CO₂ by respiration when performing light office duties [6]. Younger people such as infants and primary school children have lower emission rates but, they are likely to be more active and may well have CO₂ production at similar levels [29].

Normally, in buildings, CO₂ concentrations below 0.1% (1,000 ppm) are required to avoid discomfort and headaches [31]. According to the Canadian Center for Occupational Health and Safety [32], and also the ASHRAE standard [17], health effects can become acute at higher exposure levels.

Increased CO₂ content of the external atmosphere causes the decreased pH value of the blood [33]. Metabolism is very sensitive to body fluid pH value. Health concerns is another reason, besides global warming why the atmosphere CO₂ concentration should be limited to <426 ppm [27].

Volatile Organic Compounds (VOCs) Thousands of chemical compounds have been identified in the indoor environment. The most common pollutants are given in [17, 34]. Contaminants such as formaldehyde, toluene, volatile organic compounds (VOCs), allergens, and radon can accumulate in poorly ventilated buildings, causing health problems.

Complex mixtures of organic chemicals in indoor air also have the potential to invoke subtle effects on the central and peripheral nervous system, leading to changes in behavior and performance [35]. The latest overview on knowledge and research outcomes concerning the relationships between indoor air

pollutants and health effects highlighted future research directions [36]:

- Development of validated measuring methods
- Establishment of dose–response relationships
- Development of risk indicators for multiple exposures

Natural Ventilation for Cooling

To obtain cooling and remove heat from internal spaces, the incoming air from surroundings must be cooler than indoor temperature. The potential sources for the cool air may be from a shaded or landscaped space or from over a body of water, a labyrinth with heavy thermal mass, underground channels, or other source of cooling.

Cross ventilation is normally the primary strategy for passive cooling. Solar chimneys can be constructed to capture solar radiation to increase the difference in temperature between incoming and out-flowing air to enhance stack ventilation. In moderate and cold climates, nighttime ventilation can be applied to passively cool the building structure and provide a heat sink during the daytime occupancy period to achieve good thermal comfort.

Openable windows, as the most commonly used vents in natural ventilation systems. There are mainly four groups windows, e.g., sliding (sash), horizontal-vane opening, vertical-vane opening, and tilt and turn windows, as illustrated in a summary table in BSRIA guide [37].

Windows selection, integrated with building form and orientation, façade details, and internal layout design contribute to create different indoor airflow patterns and provide different options for the direction and volumetric flow control. In certain conditions, i.e., local drafts, cold radiation in winter, or solar gain in summer, windows can cause localized discomfort. However, occupants of naturally ventilated buildings are generally willing to accept a wider range of internal temperature and prefer more control over their environment.

The design of traditional Mosque in Malacca, Malaysia (Fig. 14) demonstrates the combination of cross ventilation, stack or heat stratification induced ventilation to achieve cooling with natural ventilation.

Daylighting

Daylight is good for health and saving energy. Natural light has a balanced spectrum of colors and wavelengths which vary over the day depending on latitude and seasons (Fig. 15). Studies suggest that daylight has benefits over artificial light sources in regulating circadian rhythms and maintaining overall health [38].

The daylight penetration depends on the room geometry as does the air distribution. Dark colors make the space feel smaller and more enclosed, whereas lighter colors have the opposite effect. In order to create stimulating high quality interior environments, lighting design must consider source intensities, distribution, glare, color rendering, and surface modeling [39]. Improved daylight metrics can be applied in a practical, real-world context to take into account for the temporal and spatial aspects of daylight, as well as meeting design standards for energy and occupants comfort [40].

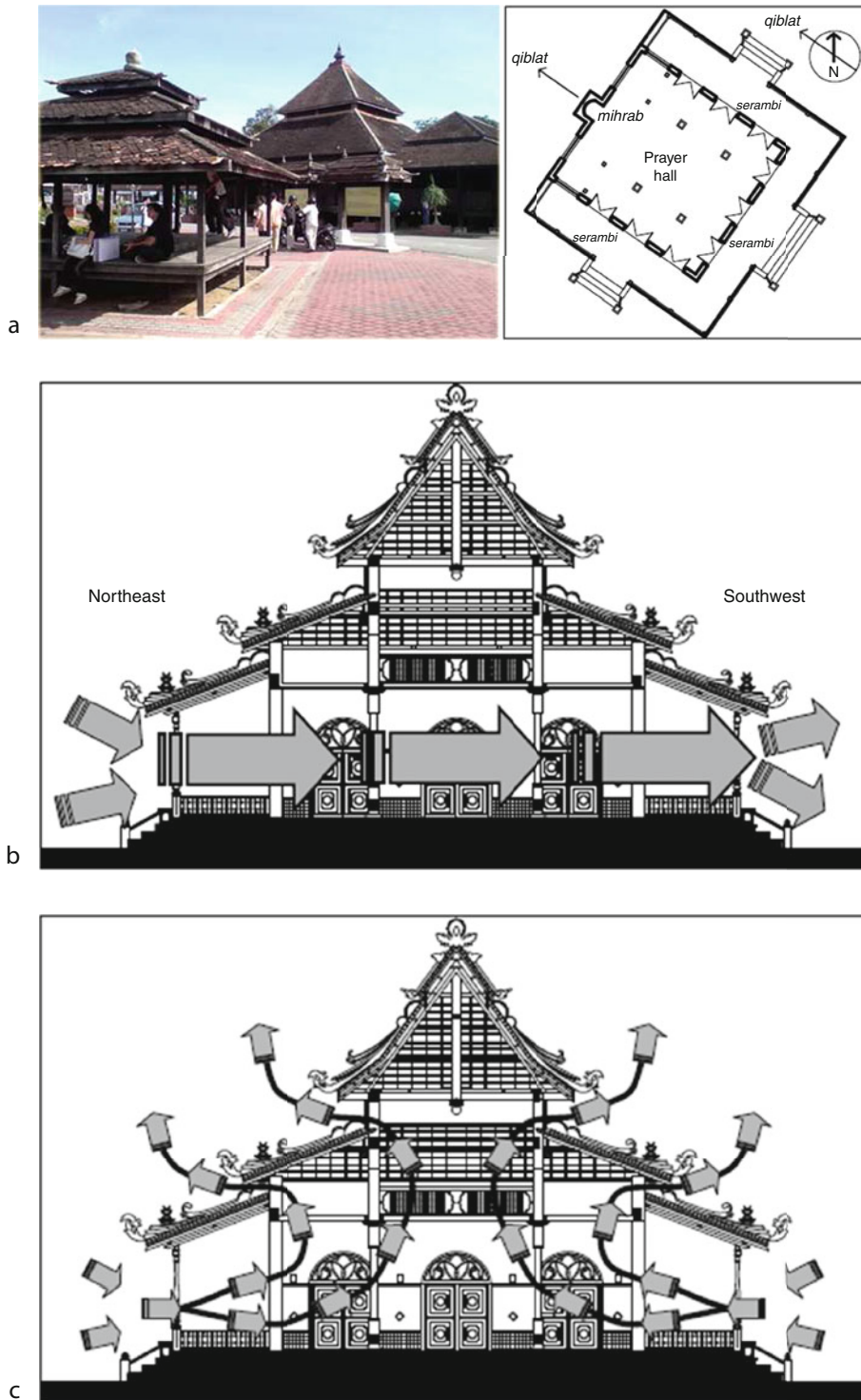
Urban Pollution, Noise, and Natural Ventilation

In the build-up urban environment, buildings and roads make up the basic geometric form of street canyons. Similar to a natural canyon, which is a steep gorge with very high sides and a minimal valley floor, an urban canyon has narrow street space bordered by very high buildings. Example of an urban canyon is the Magnificent Mile in Chicago as shown in the picture (Fig. 16).

An urban boundary layer rises above the canopy (see Fig. 17). The potential for natural ventilation is seriously affected by the reduction of wind speed, complicated turbulent dispersion patterns, elevated day- and nighttime ambient temperatures due to the urban heat island effect and increased external pollutants as well as noise level.

Based on neural network methodologies, an algorithm calculating the optimum sizes of openings for naturally ventilated buildings located in urban canyons for single-sided and stack-effect configurations was derived by Ghiaus and Allard [43].

Mitigation of the urban heat island effect can be accomplished through the use of green roofs and the use of lighter-colored surfaces in urban areas, which reflect more sunlight and absorb less heat. Green roofs protect the roof materials from intense solar radiation



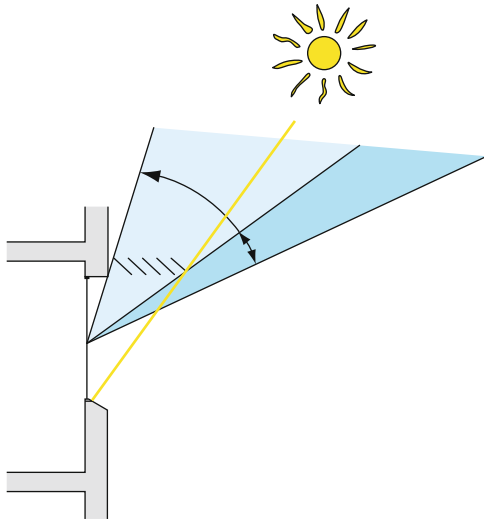
Natural Ventilation in Built Environment. Figure 14

Natural cooling in traditional Malacca Mosque, Malaysia [Source: Professor Vivian Loftness at Carnegie Mellon University]

(a) A traditional Malacca Mosque in Malaysia and typical floor plan; (b) Cross ventilation design analysis, viewing toward Southeast facade; (c) Stack-effect design analysis

and prolong the service life time. Plants retain and absorb rain improve the microclimate and also reduce the runoff water to drainage systems.

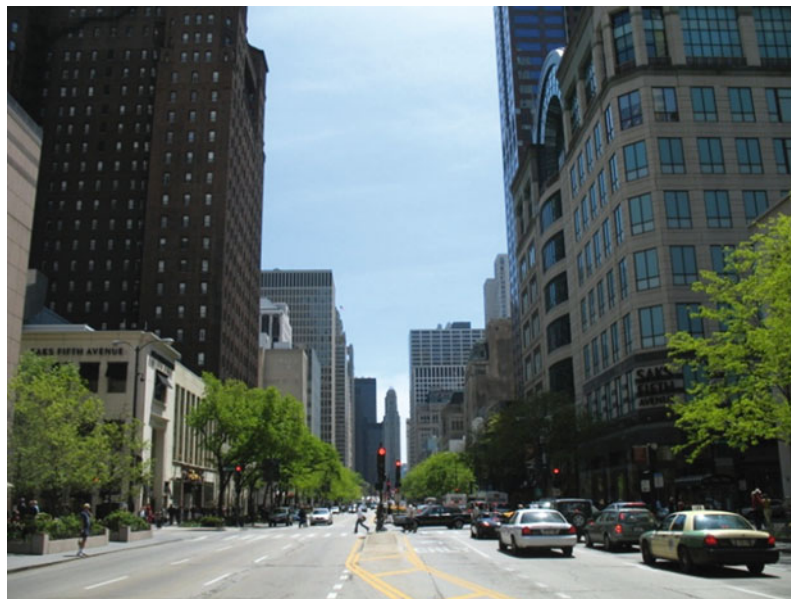
Close to nature, even in urban settings has been long-term inspiration and challenge for architects and



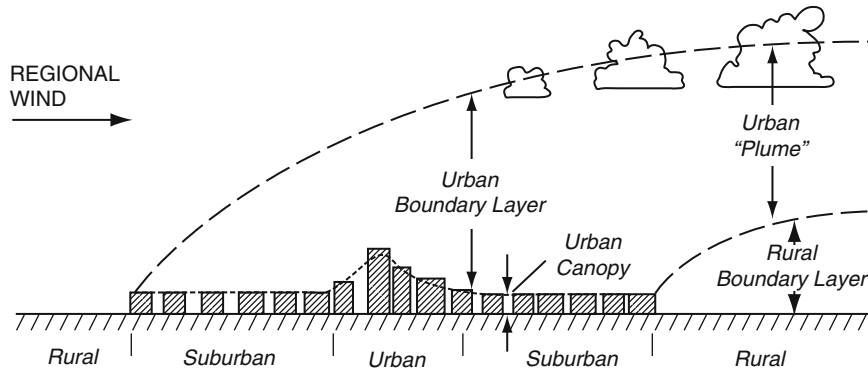
Natural Ventilation in Built Environment. Figure 15
Design to maximize daylight throughout the year [39]

engineers. In Japan, Osaka Gas Corporation sponsored an experimental “Open Building” (Fig. 18) project NEXT21 since 1994. The structural and building services have same basic units using resources more effectively through systemized construction [44]. A variety of residential units have been designed by different architects’ practices to accommodate varying households. Substantial natural greenery was planted on the “3D streets” formed by different levels of building service pathways connecting different apartments in a high-rise structure. Energy efficient measures included fuel cells and also encouraging occupants to become more aware of how to lead a comfortable life possible without increasing energy consumption.

An early study in climate chamber revealed that a change of 2.4 decipol in the PAQ or a change of 3.9 dB in the noise level has the same effect on thermal comfort as 1°C change in the operative temperature [46]. Recently, a number of newly built schools in the UK with different ventilation strategies have shown that the complex interaction between thermal comfort, ventilation, and acoustics are major challenges for designers [47].



Natural Ventilation in Built Environment. Figure 16
Urban canyon – the magnificent Mile in Chicago [41]



Natural Ventilation in Built Environment. Figure 17

Cross-section of the urban atmosphere [42]



Natural Ventilation in Built Environment. Figure 18

Open building in Japan [45]

The shape of the room and finishing materials of the surfaces affect the sound distribution. For example, concert halls are densely occupied and for music demand high spaces with volumes of 10 m^3 per person. The optimum balance of direct and indirect sound depends on the shape of the space and the boundary surface sound absorption. Combined acoustic and airflow

design chart and equations could help designers to achieve both adequate acoustical insulation and airflow rates requirement, especially in the early stages of the design process [48]. The newly completed broadcast center in London has showcased all kinds of sustainable technologies to achieve the world's first naturally ventilated television studios (see detailed in case study).

Humidity and Condensation Management with Natural Ventilation

Ventilation Effectiveness Ventilation effectiveness is an indicator of how efficiently supplied fresh air is mixed and distributed in the occupied space, it related to both the dilution and removal of indoor airborne contaminants [49]. Gan [50] used air flow pattern, air temperature, and local mean age of air (i.e., the average time for air to travel from an inlet to any point in a room and is equal to the room volume divided by the air supply rate) to determine the effective depth of fresh air distribution in a naturally ventilated space. CFD predictions showed that the width and height of window opening, room heat gains, and outdoor air temperature have combined effects in determining the maximum room depth for effective fresh air distribution in single-sided natural ventilation. For summer cooling requirement, thermal comfort is the determine factor compared to indoor air quality. Different from passive contaminants concentration based effectiveness measures [20], Coffey and Hunt [51] proposed three measures of ventilation effectiveness based on the active buoyancy (e.g., the heat or coolth) removal applied on natural displacement and natural mixing flows within a space.

Practical design guidance for naturally ventilated performing arts buildings in an urban context has been outlined in Short and Cook (2005) [52]. Specific space features and operating requirements in designing auditoria were addressed and demonstrated through the presentation of three case studies. The additional technical challenges compared to common natural ventilation space design include:

- Sizing large inlet and outlet areas
- Managing acoustic attenuation
- Configuring building management system to cater for all levels of occupancy density
- Ensuring the stratification of warm, stale air forms above the breathing zone in theaters with raked seating
- Avoiding airflow imbalance generated by wind pressure

In hospital environment, natural ventilation system design need to pay special attention to eliminate the spread of biological (i.e., fungi, bacteria, and virus), chemical, and other contaminants effectively [53].

Humidity and Condensation Prevention Low ventilation rates may lead to high indoor relative humidity. Penetration of rainwater or snowmelt into the building envelope can also cause moisture accumulation into building structures or materials. Damp structures can lead to mold formation and indoor air contamination which has been shown to be strongly associated with adverse health outcomes [54].

Low ventilation rates and moisture accumulation may lead to increased dust mites in residential dwellings; the house dust mite allergen causes asthma. Mold is also associated with exacerbation of asthma and upper respiratory disease in both children and adults [3]. In Nordic residential buildings, monitoring data has shown health risk for the residents with ventilation rates below 0.5 ACH.

Studies of ventilation rate and health effects in public buildings [55] indicate that ventilation rates below 10 l/s per person have significantly associated with health risks and PAQ complaints, increased ventilation rate between 10 and ~20 l/s per person reduce sick building syndromes (SBS) and improve PAQ.

A relative humidity range of 40–60% is generally acceptable. High humidities over 60% gradually increase the risk of mold growth, and other fungal contamination, which may cause allergy and malodors. Increased humidity may also enhance other emissions in buildings, e.g., formaldehyde from furnishings [56]. A low humidity (<30%) may cause dryness and irritation of skin, eyes, and airways of some occupants [29] leading to increased throat infections. Contact lens wearers often experience discomfort in dry environments.

Correct selection of insulation materials in modern more air tight buildings will help to develop low energy buildings in refurbishment and new design but it is important to incorporate ventilation means such as trickle ventilators for example.

Fire Safety

In the event of a fire, smoke can follow natural ventilation paths, natural ventilation system must integrate fire safety strategy and provide solution which facilitates safe occupancy, escape or increased visibility for the fire service [6]. Smoke ventilation designs utilize the buoyancy of hot smoke, operates by opening

automatic opening vents (AOVs) on the fire floor in conjunction with an AOV at the top of the smoke ventilation system to naturally extract smoke from the common escape routes. Depending on the building envelope and structures AOVs should be placed in natural/mechanical smoke shaft, atria, internal glazed screen/façade, escape stair/fire fighting stair, or external wall to achieve the prime objective of keeping common escape routes clear of smoke. Stand by fans should be installed as an emergency safety precaution.

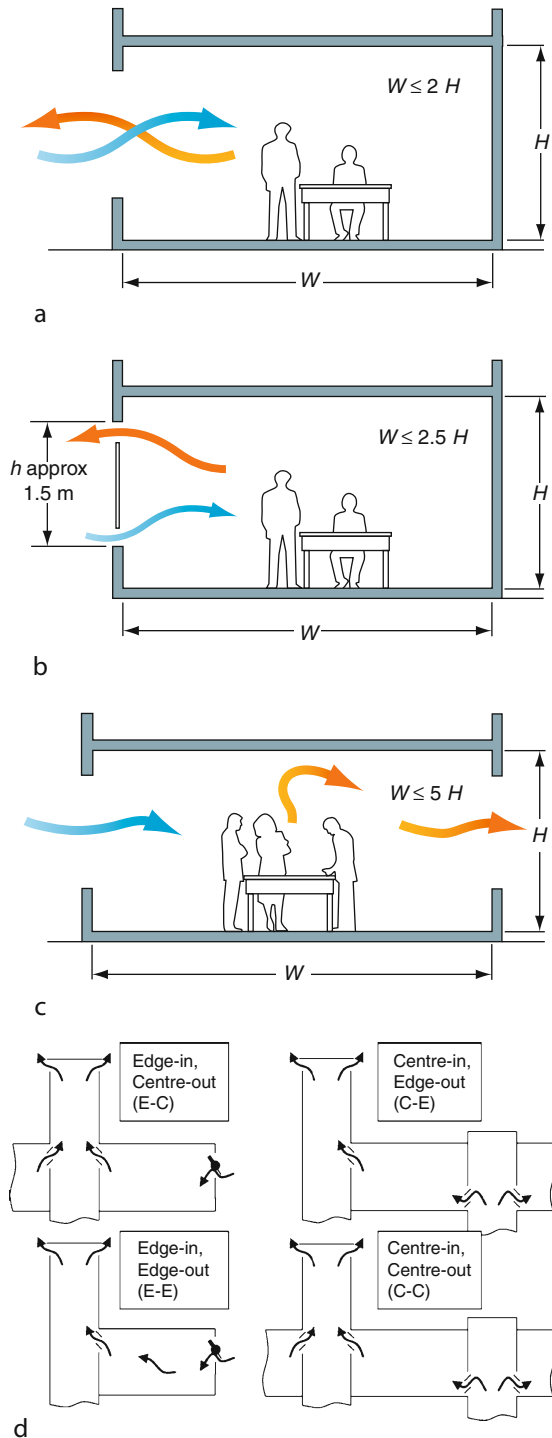
Design Guidelines

The rules of thumb for natural ventilation to be effective are as follows [10]:

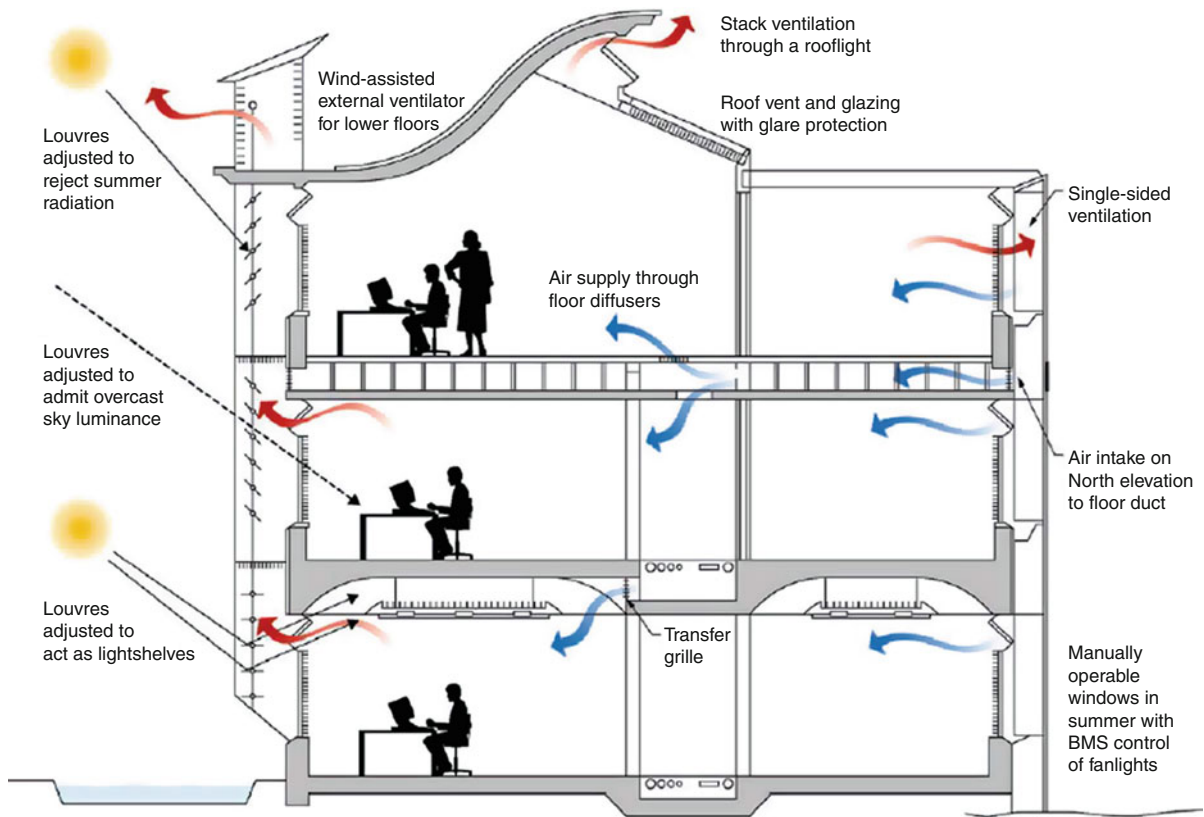
- Single-sided single-opening (mainly driven by wind turbulence) effective up to a depth of two times the floor-to-ceiling height, typically 4–6 m (Fig. 19a).
- Single-sided double-opening (mainly driven by buoyancy forces) effective up to a depth of 2.5 times the floor-to-ceiling height, typically 7–8 m (Fig. 19b).
- Cross ventilation with ventilation openings on both sides, generally opposite sides, of a space (mainly driven by wind driven) effective up to a depth of five times of the floor-to-ceiling height, typically 15 m (Fig. 19c).
- Stack ventilation is mainly driven by temperature differences between the hot air in the occupied space and the cooler external air. The effective depth of stack ventilation is up to five times of floor-to-ceiling height. Stack ventilation can also be enhanced by wind effect or through the use of a solar chimney, i.e., solar driven stack ventilation [57].
- Different forms of atrium ventilation are illustrated in Fig. 19d [58].

Figure 20 illustrates various natural ventilation strategies can be integrated into design.

For natural ventilation aim for a heat gain less than 35 W/m^2 to avoid excessive overheating. This means there is a need to reduce internal gains; ensure effective ventilation; select suitable facades and materials. Also provide CO_2 monitors to help occupants know when it is preferable to increase ventilation.



Natural Ventilation in Built Environment. Figure 19 Schematic diagrams of the different forms of natural ventilation (a) single-sided single-opening; (b) single-sided double-opening; (c) cross ventilation (d) Atrium ventilation [58]



Natural Ventilation in Built Environment. Figure 20
Illustration of various natural ventilation strategies [37]

Passive cooling can be achieved by using:

- Local climate characteristics such as breezes
- Intelligent facades
- Solar protection
- Low energy lighting
- Thermal mass
- Ventilation
- Cloud computing, which can reduce computer heat gains considerably

Evaporative down draft cooling technique [59, 60] involves introducing fresh ambient air at the top of a central lightwell and cooling it so that it flows downward, filling the space with a static reservoir of denser, cooler air.

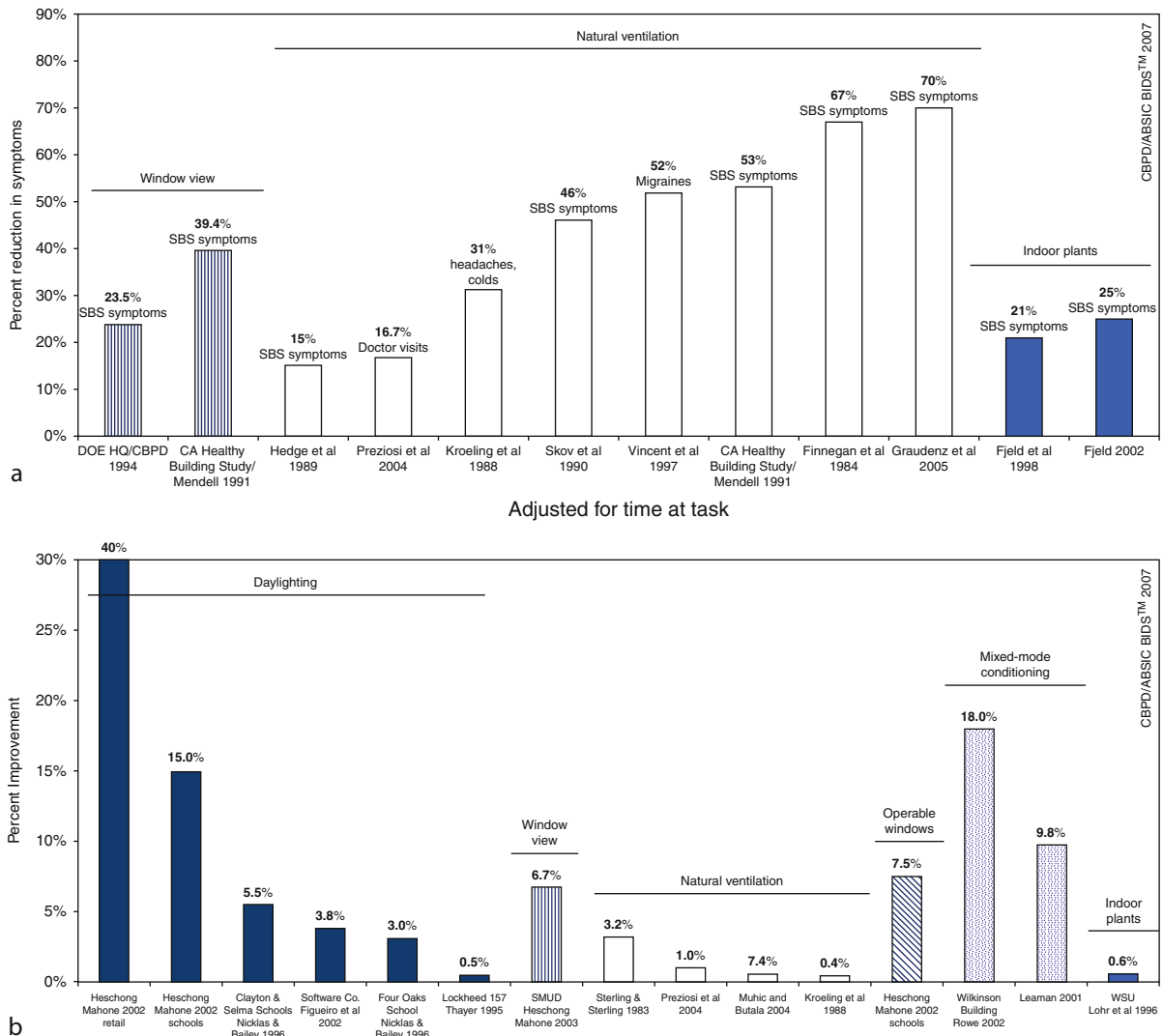
Benefits of Natural Ventilation

The parameters which affect the air velocity and temperature at a given point in the room are as follows:

1. Air inlet velocity; sound emission must also be accounted for when selecting a value for this
2. Supply to room temperature differential
3. Geometry and position of air supply outlet
4. Position of air extract outlet
5. Room geometry
6. Room surface temperatures; low surface temperature components, such as glass, tend to promote strong convection currents
7. Position, shape distribution, and emission of heat sources (e.g., people)
8. Room turbulence

The advantages of natural ventilation arise because there are no mechanical systems hence:

- Less energy is consumed
- Less plant room space is needed
- Higher level of daylight if well-designed
- Maintenance is simpler



Natural Ventilation in Built Environment. Figure 21

Health and productivity gains from access to the natural environment [61] (a) Health gains from access to the natural environment; (b) Productivity gains from access to the natural environment

- Increased durability
- No noise
- Good occupant control using windows
- Low cost

The disadvantages are as follows:

- There is no filtration or control of moisture content.
- The driving pressure depends on the wind and/or the stack effect and both are variable and cannot be easily be controlled.

- There has to be an integrated approach to design between the architect and the engineer with regards to built form; orientation; massing; internal layout; selection of window types and their positioning in the façade.
- Internal heat gains are limited to less than 35 W/m².

The ventilation rate depends on the strength and direction of wind and/or buoyancy forces and the resistance of the flow path. The uncontrollable feature of natural ventilation can result in the air change rate

varying significantly and being distributed unevenly to internal spaces giving periods of inadequate ventilation, or periods of over ventilation and excessive energy waste.

Health and Productivity

There is a unique relationship between an individual, the environment, and the building they inhabit. The complicated interaction between the “hard” environmental stimuli, such as air and surface temperatures, humidity, air movement, and air purity, and the “soft” interlinked social and psychological factors of individuals and their organization, influences the sense of well-being, consequently the productivity. Task performance is best when the mind is alert at an optimum arousal level with the least distraction [27].

Loftness [61] shows the impact of natural ventilation on health and productivity in the Fig. 18a and b. In addition to the health and productivity benefits brought by design with access to the natural environment, effective daylighting can yield 10–60% reductions in annual lighting energy consumption. There is evidence of potential 40–75% reductions for cooling energy consumption when natural ventilation is inter-actively supported by mixed-mode HVAC systems.

Sustainable and healthy built environments result from integrating the natural diversity of the region – its unique climate and seasons, textures, sounds, smells, and variety of landscape and species.

Selection of Ventilation Strategies

The CIBSE AM 10 1997 [62] gives the following monograph (Fig. 22) to help the decision process for selecting whether to use natural ventilation, mechanical systems, or air-conditioning.

Design Elements for Natural Ventilation

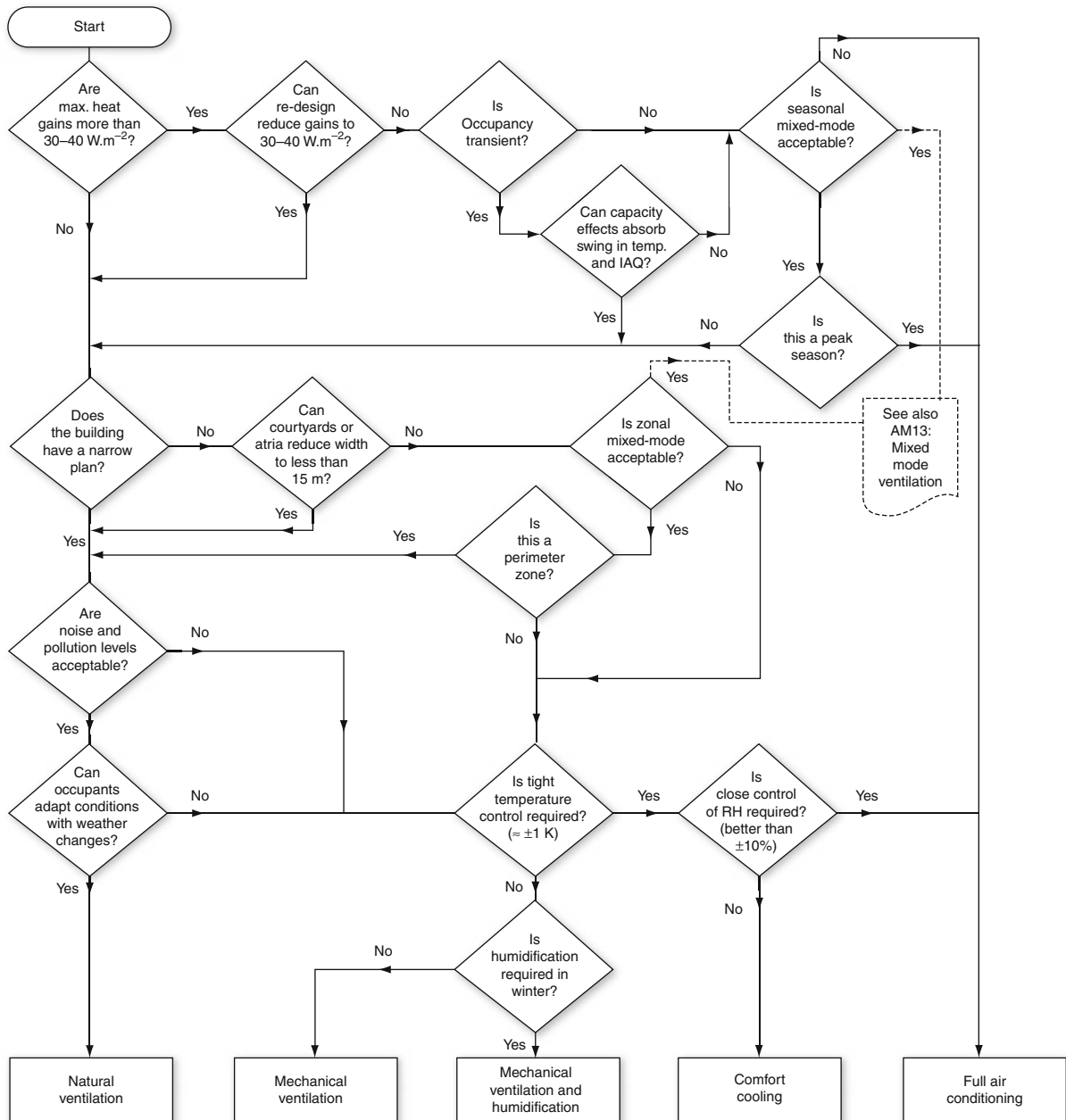
Natural ventilation can be difficult to control due to the fluctuating indoor and outdoor conditions. As previously stated naturally ventilated buildings have to be inextricably linked to architectural form and fabric; they require holistic design and significant attention to detail.

Well-designed natural ventilation systems need to address the following aspects comprehensively [62–64]:

- *Site design* – building location, orientation, site layout, and landscaping
- *Building design* – building type and function, building form and orientation, envelope, thermal mass, natural ventilation strategy, internal spatial division and functions, internal heat load, solar shading, daylight, passive night cooling potential
- *Vent opening design* – position of openings, clear path of airflow, types of openings, sizing and choice of window opening design, effective area of multiple openings, provision of secure, operable openings, and control strategy

Computer Aided Prediction Models

Chen [65] presented an overview of ventilation performance prediction methods, including analytical models, empirical models, small-scale experimental models, full-scale experimental models, multizone network models, zonal models and Computational Fluid Dynamics (CFD) models. Recent applications of the above simulation tools were also examined in terms of contributions for practical design and research purposes. CFD applications in modeling of wind-driven natural ventilation [66] have shown improved prediction capability for complex naturally ventilated buildings. Walker [67] developed a methodology to evaluate natural ventilation in a multizoned commercial office building by full-scale building monitoring, reduced-scale physical experiment and CFD simulation. Detailed human thermal sensation and comfort models provide more accurate predictions on the dynamics responses of occupants to building environments, also advanced coupling simulation system extends the prediction capability of CFD and brings the human occupancy factor further into the core activities of the design process of buildings [68, 69]. The integration of CFD with dynamic building simulation (BS) models and Geographic information systems (GIS) data would be a practical way to take advantage of the strength of the other models for optimal natural ventilation design and analysis. Figure 23 illustrates CFD application on advanced natural ventilation design in hospital. Lomas and Ji [70] evaluated simple natural ventilation (SNV) and advanced natural ventilation (ANV) design in terms of overheating risk in the current and future



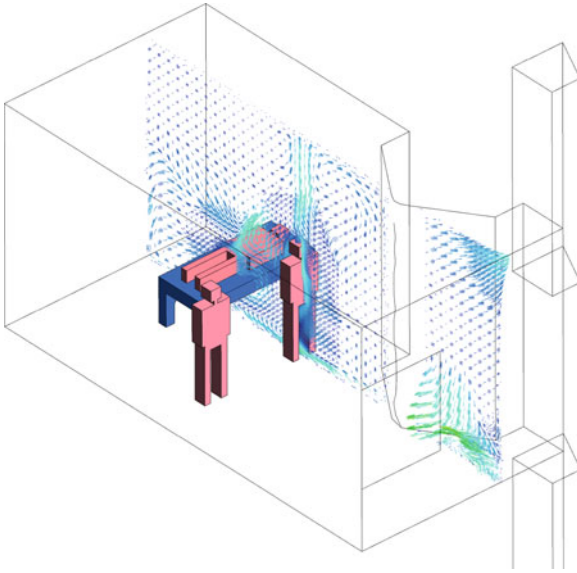
Natural Ventilation in Built Environment. Figure 22

Flowchart for selecting a ventilation strategy [62]

climate in health care buildings. They also proposed an overheating risk criterion compatible with adaptive thermal comfort assessment. Both field monitoring and modeling studies showed that ANV could offer

greater resilience to climate change than SNV, particularly as a refurbishment strategy.

The strengths and weaknesses of a wide range of design simulation tools are summarized by



Natural Ventilation in Built Environment. Figure 23
Advanced natural ventilation system design in hospital
using CFD [45]

governmental and professional organizations [71, 72]. Educational resources and practical equations for various design stages and step-by-step guided case studies [58, 73–75] would help multidisciplinary professions to design and build sustainable buildings.

The following case studies showcase innovative solution to natural and mixed-mode ventilated buildings in the built environment.

Case Studies

Case Study 1: Liverpool John Moores University, UK

The Liverpool John Moores University art and design academy was designed by award winning Rick Mather Architects and engineered by Whitbybird Engineers and was built by Wates Construction. The building was designed from 2004 and opened in 2008 (see Fig. 24).

The 11,000 m² five story university academy building had a number of important environmental initiatives on the client brief including a 25% betterment of UK conservation of fuel and power building regulations; a BREEAM target of Very Good, and a 10% onsite renewable energy target.



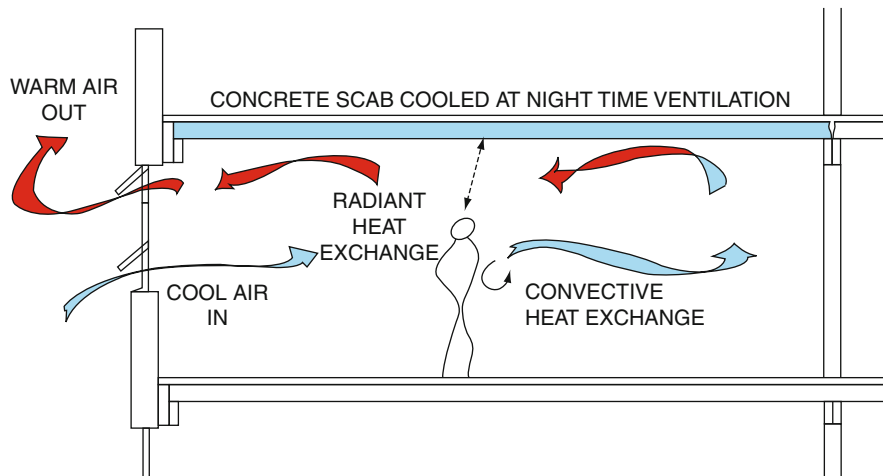
a



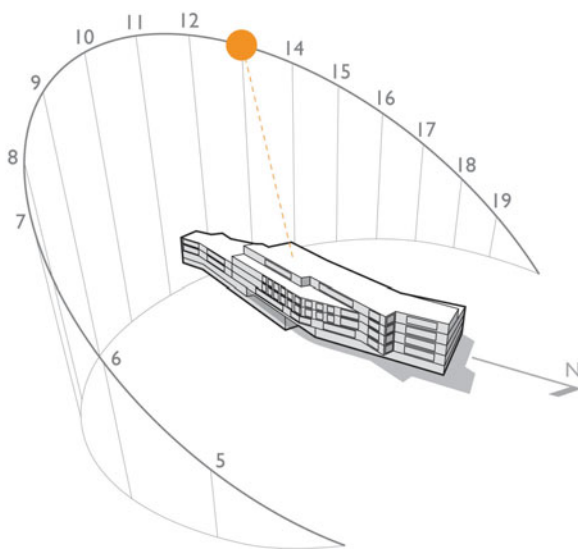
b

Natural Ventilation in Built Environment. Figure 24
LJM University Liverpool façade in nighttime (a) LJM University
Liverpool exterior; (b) LJM University Liverpool façade in nighttime (Source:
Rick Mather Architects)

To meet these requirements the design team reviewed a number of architectural and structural and services engineering options and concluded that a mixed-mode ventilation scheme utilizing thermal mass with Heat Recovery would of particular benefit to a low energy approach. The building design team used the latest computer modeling techniques to prove compliance with industry standards for thermal



Natural Ventilation in Built Environment. Figure 25
Exposed concrete soffit for thermal mass benefit



Natural Ventilation in Built Environment. Figure 26
Sun path analysis for external façade design (Source: Whitbybird Engineers)

comfort particularly the prevention of overheating (see Figs. 25, 26, 27, and 28).

The building benefits from a heavyweight thermal mass during the spring and summer, utilizes single-sided natural ventilation with mechanical extract during peak summer with nighttime free cooling (see Fig. 25), and during winter operates as a sealed building with heating provided by a biomass pellet boiler.

A weather station provides data to the BEMS to ensure the automated windows are aligned with the ventilation strategy.

The building was shortlisted for various awards in 2010.

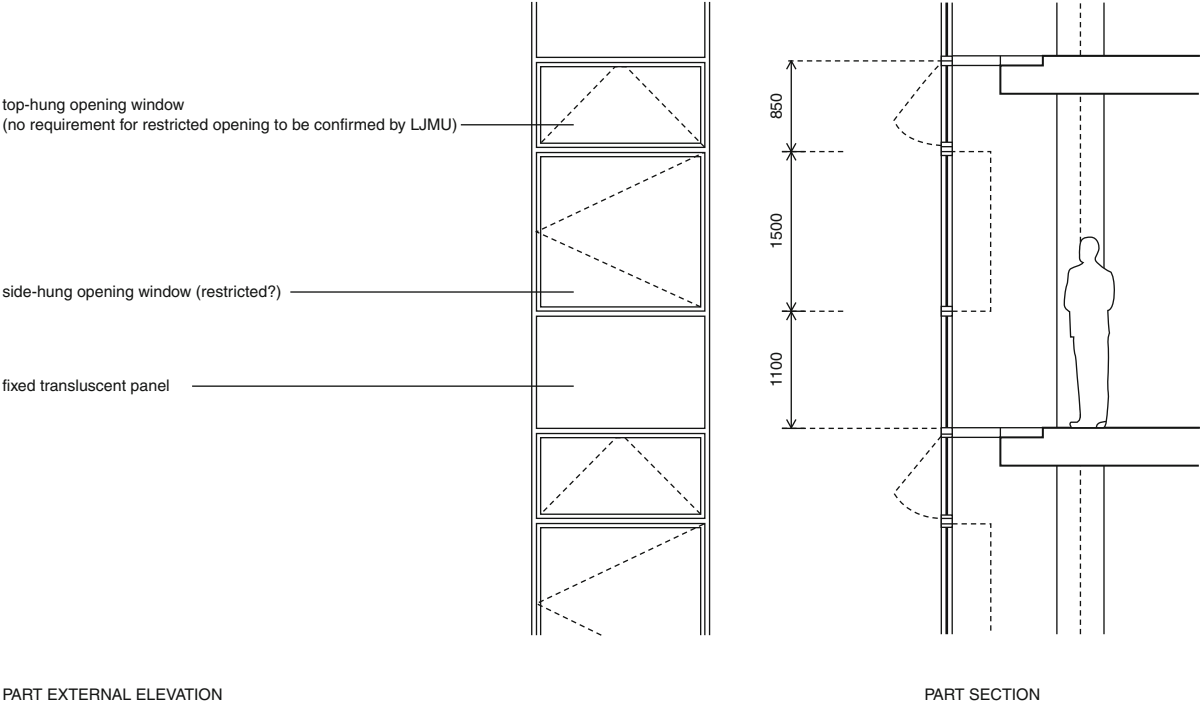
Case Study 2: Tamworth Academies – Staffordshire, UK

Two secondary school academies were commissioned by Tamworth County Council in 2009. Designed by Aedas architects and engineered by WSP and CTM the buildings were designed in 2009/2010 and are scheduled to open in 2010. Main contractors are Wilmott Dixon Construction.

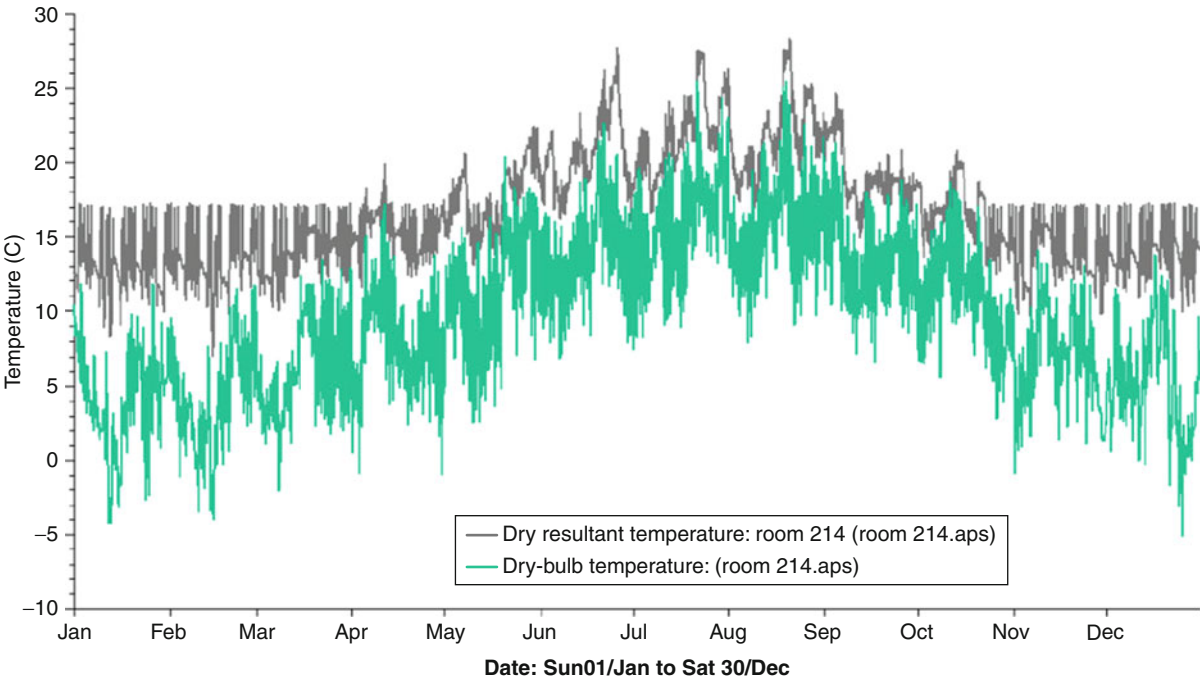
The two 9,000 m² two story buildings (Fig. 29) had to comply with the government's 60% carbon reduction target compared to 2002 UK building regulations. The design team also set a BREEAM environmental target of Excellent.

It was the aim of the architectural and structural and services engineering teams to strive to design a low energy building to ensure the building was future proofed against rising energy costs. In addition the client team wanted a low whole life cost solution.

The selected structural and services strategy was a building that would be designed to operate as a naturally ventilated building but be capable of operating as a highly efficient mechanically ventilated



Natural Ventilation in Built Environment. Figure 27
External façade design (Source: Rick Mather Architects)

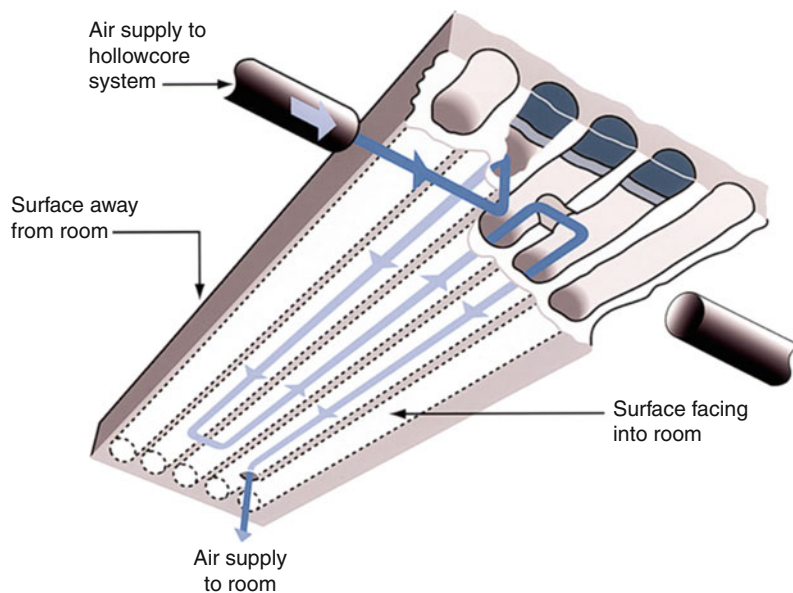


Natural Ventilation in Built Environment. Figure 28
Thermal modeling chart (Source: Whitbybird Engineers)



Natural Ventilation in Built Environment. Figure 29

Tamworth Landau Forte exterior (Source: Professor Vivian Loftness at Carnegie Mellon University)

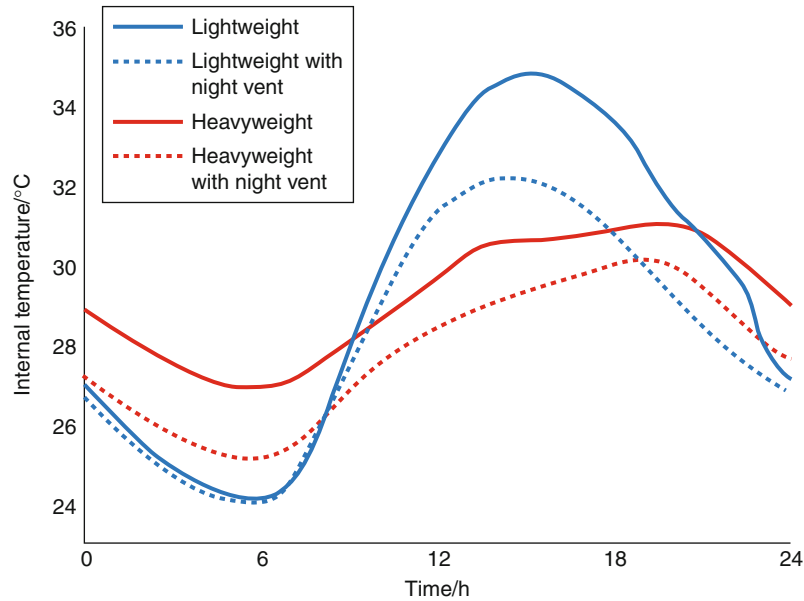


Natural Ventilation in Built Environment. Figure 30

Termodeck concrete plank (Source: Tarmac Termodeck)

building during peak summer and extremely cold winters to conserve energy. The client, main contractor, and design team chose Termodeck by Tarmac, an integrated structural mixed-mode ventilation scheme utilizing thermal mass free cooling with energy efficient heat recovery (see [Figs. 30](#) and [31](#)).

Other features of the buildings include areas of the academies where ICT gains are increased provided with chilled water generated by a ground source heat pump. And seasonal operation of the ventilation strategy communicated to the occupants of the academies via building weather station data and LCD flat screen TVs.



Natural Ventilation in Built Environment. Figure 31

Temperature time-lag for lightweight and heavyweight buildings (Source: Tarmac Termodeck)

Case Study 3: BSKyB Broadcast Center – London, UK

Designed by Arup Associates, the BSKyB broadcast center (Fig. 32) located in west London houses the world first naturally ventilated television studios [76]. Thirteen giant ventilation chimneys, nine line the building's eastern elevation and another four appear on the west façade. Concrete boxes within boxes construction provides solution to eliminate external noise as well as naturally ventilate the studios to remove excessive heat generated by studio lights. Fresh air is supplied through acoustically lined labyrinth built in between the underside of the studio's concrete floor and the floor of the surrounding box from street level (see Fig. 33). This construction form allows big air paths to minimize resistance to air movement as well as eliminate all influx of noise.

In order to prevent a common stack ventilation problem – air cooling in the flue and dropping back into a room – the flues are lined and insulated on the inside. In an intermediate mode the ventilation system will run on extract only to pull the air up the chimney and warm it. When the right flue surface temperature has reached, the air's natural buoyancy will take over, the system then switch to natural ventilation mode automatically.



Natural Ventilation in Built Environment. Figure 32
BSKyB broadcast center in west London [76]



Natural Ventilation in Built Environment. Figure 33
Natural ventilation flow within the BSkYB television studios
(Source: Arup Associates)

At the south end of the building, a glazed atrium houses a series of meeting rooms, a café, and breakout spaces, which allows access between levels. The office areas (8 m in depth) on the west elevation are ventilated using single-side natural ventilation. The offices (15 m deep) on the eastern side utilize three atrium-line chimneys in the center of the building to help draw air across the floor plates. Natural light through these atriums are additional benefit of the design.

Case Study 4: Commerzbank – Frankfurt, Germany

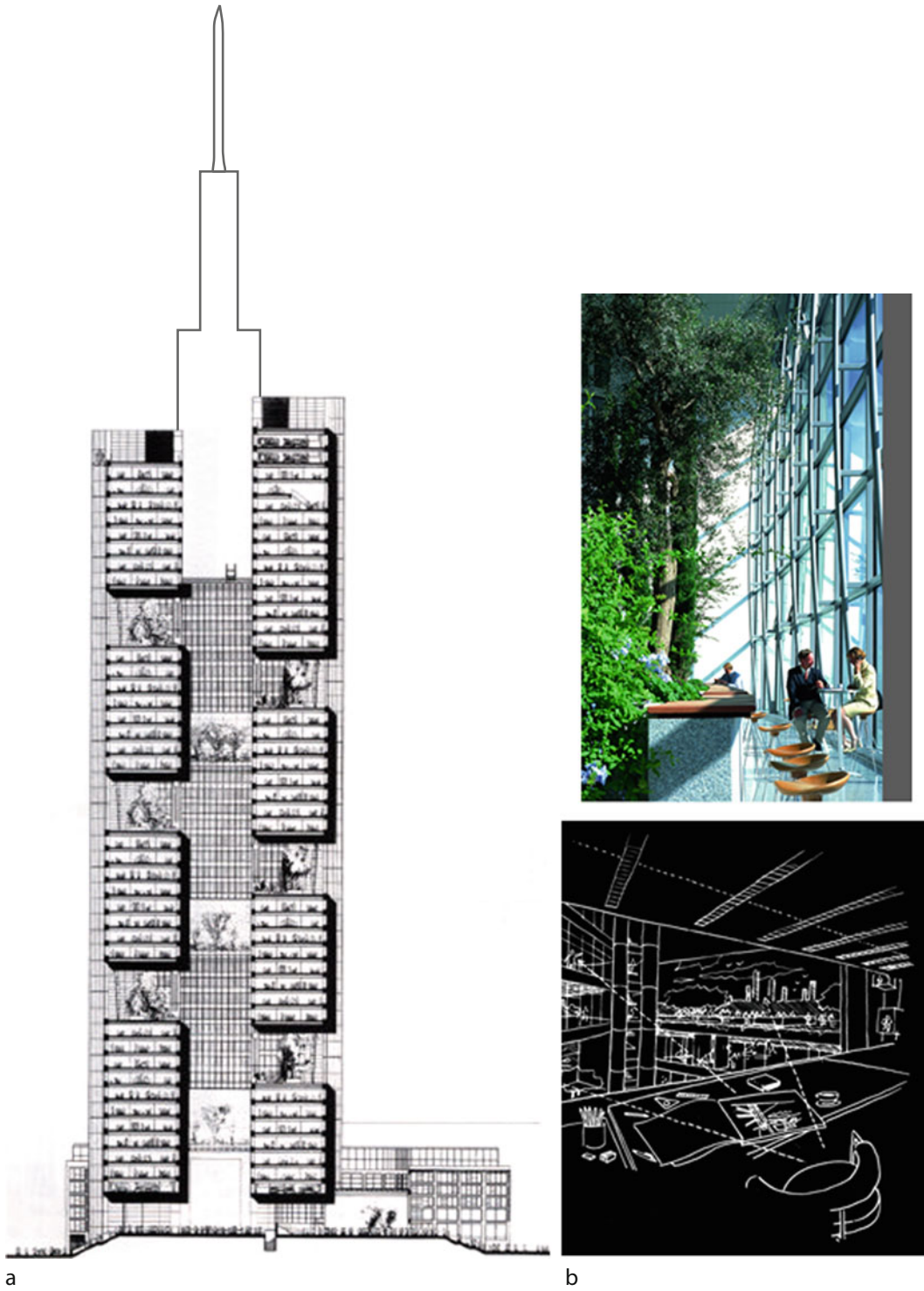
Norman Foster's Commerzbank in Frankfurt, Germany (see Fig. 34), demonstrates sustainable urban architecture featured by natural ventilation, vast amount of daylight, and pleasant exterior views within deep building can be achieved at the scale of the skyscraper.

Triangular-shaped plan provide the rigid structural support with high-rise building functional cores located at each corner of this triangle. The center atrium of the buildings provides light both vertically, from the glass roof at the atrium's top, horizontally from the winter garden facades to the office areas. Winter gardens rotate around the façade of the building, allow for ventilation through the atrium being divided into sections. Natural light are brought directly into the center of the buildings, offices facing the center are provided light and outdoor views through these green, natural spaces. The operable layered façade allows natural ventilation through office spaces, while winter gardens provide ventilation for the entire building.

Case Study 5: Queen's Building, De Montfort University, UK

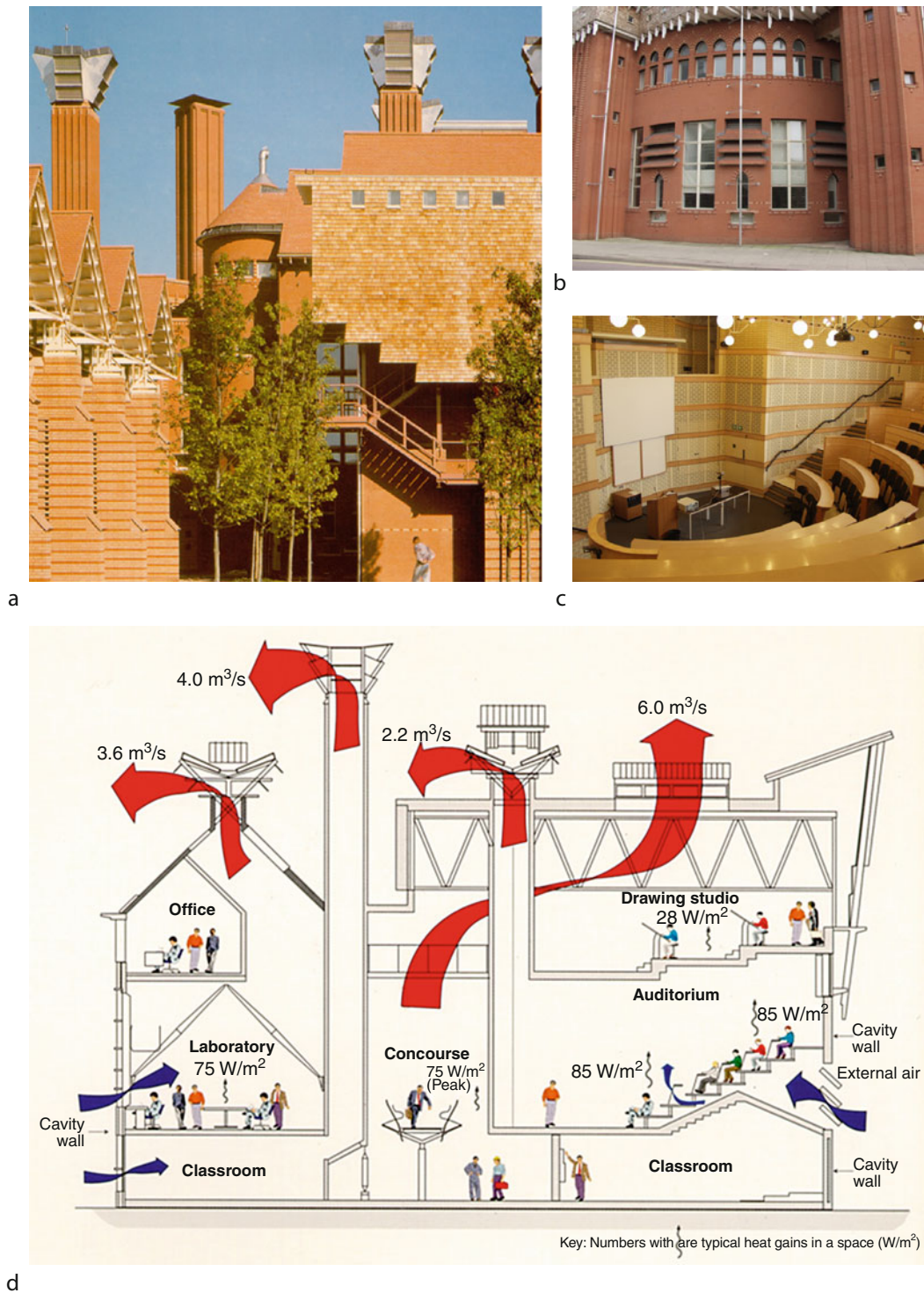
The Queen's Building (Fig. 35) designed by Short Ford Associates and built in 1993 featured with large venting chimneys, heavy thermal mass, shallow floor plan, operable windows, and generous ceiling heights to facilitate natural ventilation and daylighting. This traditional brick building has wide insulation-filled cavity walls and concrete slabs in the ceilings, buffering the indoor from outdoor temperature peaks. Glazed ventilators also help to provide as much natural lighting as possible. In the auditoria fresh air enters through louvers in the north façade by means of plenums below the raked wooden floor and wall inlets which are controlled by the BEMS.

Since being the Green Building winner of the Year 1995, Queen's building has served a great demonstration value for environmental design. It aims to a "Living Lab" to showcase innovative technologies and demonstrate ways of achieving significant carbon reductions in modern buildings through the refurbishment.



Natural Ventilation in Built Environment. Figure 34

Commerzbank, Frankfurt, Germany (a) Building section [77] (b) Winter garden and interior view design [78]



Natural Ventilation in Built Environment. Figure 35

The Queen's building, De Montfort University, Leicester, UK **(a)** Exterior view and stack outlets; **(b)** Exterior view and air inlets; **(c)** Interior view of auditorium; **(d)** The Queen's building – natural ventilation strategy

Future Directions

Sustainable design requires a long-term durable approach and passive environmental control offers this. Holistic design and construction is necessary to achieve the sustainable built environment [79, 80], provide the optimum cost-benefit value for all stakeholders in the built environment industry. Natural ventilation should be explored as a possibility for all projects, but various factors can limit its use although, throughout history, vernacular architecture demonstrates it can work effectively in many climates but only if there is a unity of thought between architects and the engineers. However hybrid systems are common in Europe emphasized by the demand for low carbon buildings. Ultimately buildings are for people and systems must provide conditions for enhancing well-being at a time when climates worldwide are showing more variability and change.

Acknowledgments

The authors would like to thank Lee Hargreaves (WSP, UK Ltd.) for his help with researching and organizing some case studies; Mike Beaven of Arup Associates for some case studies; Matt Kitson (Hilson Moran) for providing Gherkin images; Professor Vivian Loftness (Carnegie Mellon University, USA) for providing architectural images and Fig. 21; Beifan Yang and Bin Zhang (Tianjin Weland Landscape Architecture Design Co, Ltd. China) for the architectural drawings; Jin Zhang (JINT Design Consultants Ltd.) for architectural sketch; and Dr Malcolm Cook (Loughborough University) for providing the images of Queens Building.

Bibliography

- Elliott CD (1992) *Technics and architecture: the development of materials and systems for buildings*. MIT Press, Cambridge, MA
- Fathy H (1986) *Natural energy and vernacular architecture, principles and examples with reference to hot arid climates*. The University of Chicago Press, Chicago
- Miller JD (2007) Indoor air quality and occupant health in the residential built environment: future directions. In: Yoshino H (ed) *Proceedings IAQVEC 2007*, Sendai, Japan. ISBN 978-4-86163-069-9, pp 15–22
- Von Frisch K (1975) *Animal architecture*. Hutchinson, London
- Fitch JM (1976) *American building: the environmental forces that shape it*. Schocken Books, New York
- BSI (1991) BS 5925:1991 Code of practice for ventilation principles and designing for natural ventilation. BSI, London
- Texas Tech University (2004) The wind science and engineering (WISE) research center; Available from <http://www.wind.ttu.edu/>. Accessed 11 June 2004
- CIBSE (2006) Guide A: environmental design. The Chartered Institution of Building Services Engineers, London
- Liddament MW (1996) *A guide to energy efficient ventilation*. Air Infiltration and Ventilation Centre, Coventry
- CIBSE (2005) AM10: Natural ventilation in non-domestic buildings. The Chartered Institution of Building Services Engineers, London
- Liddament MW (2010) The applicability of natural ventilation. In: CIBSE Natural Ventilation Group Seminar 2010 – Natural ventilation in the urban environment. RIBA, London
- The Gherkin (2008) <http://www.30stmaryaxe.co.uk/>. Accessed 8 Sept 2011
- Abbas T (2008) MSc intelligent buildings, lecture notes. Hilson Moran, London
- Hazim A (2003) *Ventilation of buildings*, 2nd edn. Spon Press, London
- Seppänen O, Fisk W, Mendell M (1999) Association of ventilation rates and CO₂ concentrations with health and other responses in commercial and institutional buildings. *Indoor Air* 9(4):226–252
- Wargocki P, Seppanen O, Anderson J, Boerstra A, Clements-Croome D, Fitzner K, Olaf Hanssen S (2006) *Indoor climate and productivity in offices: guide book 6*. Federation of European Heating and Air-Conditioning Associations (REHVA), Brussels
- ASHRAE (2010) Standard 62.1-2010 – Ventilation for acceptable indoor air quality. American Society Heating, Refrigerating and Air Conditioning Engineers (ASHRAE), Atlanta
- ASHRAE (2009) *Handbook – Fundamentals*, chapter 9 Thermal comfort. American Society Heating, Refrigerating and Air Conditioning Engineers (ASHRAE), Atlanta
- Croome DJ, Roberts BM (1975) *Airconditioning and Ventilation of Buildings* (Pergamon Press); second edition 1981
- Etheridge D, Sandberg M (1996) *Building ventilation theory and measurement*. Wiley, Chichester
- Linke W (1956) Strömungsvorgänge in künstlich belüfteten Räumen, Forschungsberichte des Wirtschafts und Verkehrsministeriums des Landes, NRW Nr 259. *Kaltetechnik* 18:122
- Mülleijans H (1973) Über die Bedingungen von Modellversuchen in der Klimatechnik. *Ki* 8/73, Teil 6, S. 63 ff
- van Gunst E, Erkelens PJ, Coenders WPJ (1967) In 4th congress international du chauffage et de la climatisation, Paris
- Fang L, Clausen G, Fanger PO (1998) Impact of temperature and humidity on perception of indoor air quality during immediate and longer whole-body exposures. *Indoor Air* 8:276–284
- Fang L, Clausen G, Fanger PO (1998) Impact of temperature and humidity on the perception of indoor air quality. *Indoor Air* 8:80–90

26. Wargocki P (2004) Sensory pollution sources in buildings. *Indoor Air* 14:82–91
27. Clements-Croome DJ (2008) Work performance, productivity and indoor air. *Scand J Work Environ Health Suppl* 4:69–78
28. Health Canada (2010) Environmental and workplace health; Available from www.hc-sc.gc.ca. Accessed 15 Aug 2010
29. ECA (1992) Report no. 11 – Guidelines for ventilation requirements in buildings in environment and quality of life, European Collaborative Action – Indoor air quality & its impact on man, EUR 14449 EN
30. Dobson R (2008) Smoking bans reduce heart attack admissions. *British Med J* 337: a597
31. ISO:7730 (2005) Ergonomics of thermal environments – Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria. International Organization for Standardization (ISO), Geneva
32. CCOHS (2002) Health effects of carbon dioxide gas. Canadian Centre for Occupational Health and Safety, Hamilton, Ontario
33. Robertson D (2006) Health effects of increase in concentration of carbon dioxide in the atmosphere. *Curr Sci* 90(12):1607–1609
34. HSE (1990) Occupational exposure limits – guide note EH 40/90. Health and Safety Executive, HMSO, London
35. Hinds WC (1999) Aerosol technology: properties, behavior, and measurement of airborne particles. Wiley, New York, pp 464
36. Brown T, Holmes P, Harrison PTC (2010) Review: the applicability of epidemiological methods to the assessment of the risks to human health of indoor air pollution: an overview. *Indoor Built Environ* 19(3):311–326
37. Pennycook K (2009) The illustrated guide to ventilation. In BSRIA guide – BG 2/2009. The Building Services Research and Information Association (BSRIA), Bracknell
38. Boyce PR (2010) Review: the impact of light in buildings on human health. *Indoor Built Environ* 19(1):8–20
39. RAE (2010) Engineering a low carbon built environment – The discipline of Building Engineering Physics. The Royal Academy of Engineering, London
40. Mardaljevic J, Hescong L, Lee E (2009) Daylight metrics and energy savings. *Lighting Res Technol* 41(3):261–283
41. Wiki (2010) http://en.wikipedia.org/wiki/Urban_canyon. Accessed 18 July 2010
42. Barlag AB, Kuttler W (1990/1991) The significance of country breezes for urban planning. *Energy and Buildings* 15(3–4): 291–297
43. Ghiaus C, Allard F (2005) Natural ventilation in the urban environment. Earthscan, London
44. Fukao S (2010) The history of developments toward open building in Japan. Lecture at Loughborough University, UK on 16 July 2010
45. CIB (2010) W104 Open building implementation; Available from <http://www.open-building.org/ob/next21.html>. Accessed 20 July 2010
46. Clausen G, Carrick L, Fanger PO, Kim SW, Poulsen T, Rindel JH (1993) A comparative study of discomfort caused by indoor air pollution, thermal load and noise. *Indoor Air* 3:255–262
47. Mumovic D, Davies M, Ridley I, Altamirano-Medina H, Oreszczyn T (2009) A methodology for post-occupancy evaluation of ventilation rates in schools. *Building Serv Eng Res Technol* 30(2):143–152
48. Coley DA, Hunt S, Mitchell A (2009) Acoustics in schools: explaining the options to architects by the use of approximate formulae and graphs, with a special emphasis on dining spaces. *Indoor Built Environ* 18(6):505–513
49. Mathisen HM, Moser A, Nielsen PV (2004) Guidebook no. 2 – Ventilation effectiveness. In: Mundt E (ed) Federation of European heating and air-conditioning associations REHVA Journal, Brussels, Belgium
50. Gan G (2000) Effective depth of fresh air distribution in rooms with single-sided natural ventilation. *Energy and Buildings* 31(1):65–73
51. Coffey CJ, Hunt GR (2007) Ventilation effectiveness measures based on heat removal: part 2 Application to natural ventilation flows. *Building Environ* 42(6):2249–2262
52. Short CA, Cook MJ (2005) Design guidance for naturally ventilated theatres. *Building Serv Eng Res Technol* 26(3):259–270
53. WHO (2009) Natural ventilation for infection control in health-care settings. World Health Organization, Geneva
54. ASHRAE (2009) Indoor Air Quality Guide - Best Practices for Design, Construction, and Commissioning, in ASHRAE Design Guide. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta
55. Seppänen O, Fisk W (2004) Summary of human responses to ventilation. *Indoor Air* 14(7):102–118
56. ISIAQ (1996) Control of moisture problems affecting biological indoor air quality. In: Task Force Report. International Society of Indoor Air Quality and Climate (ISIAQ), Austin, TX
57. CIBSE (2005) Guide B: heating, ventilating, air conditioning and refrigeration. The Chartered Institution of Building Services Engineers, London, pp 2–9
58. Lomas KJ (2007) Architectural design of an advanced naturally ventilated building form. *Energy and Buildings* 39(2):166–181
59. Short CA, Cook MJ, Woods A (2009) Low energy ventilation and cooling within an urban heat island. *Renewable Energy* 34:2022–2029
60. Ford B, Schiano-Phan R, Francis E (2010) The architecture and engineering of downdraught cooling: a design source book. PHDC Press, London. ISBN 978-0-9565790-0-3
61. Loftness V, Snyder M (2008) Where windows become doors. In: Kellert S, Heerwagen J, Mador M (eds) Biophilic design. Wiley, Hoboken, pp 119–131
62. CIBSE (1997) AM10: natural ventilation in non-domestic buildings. The Chartered Institution of Building Services Engineers, London
63. Cohen R (1997) Environmental criteria for naturally ventilated buildings. In: Clements-Croome D (ed) Naturally ventilated buildings – Buildings for the senses, economy and society. E & FN SPON, London
64. Allard F (1998) Natural ventilation in buildings - a design handbook, James & James Ltd, London, pp. 366. ISBN 9781873936726.

65. Chen Q (2009) Ventilation performance prediction for buildings: a method overview and recent applications. *Building Environ* 44(4):848–858
66. Jiru TE, Bitsuamlak GT (2010) Application of CFD in modelling wind-induced natural ventilation of buildings – a review. *Int J Ventilation* 9(2):131–147
67. Walker CE (2006) Methodology for the evaluation of natural ventilation in buildings using a reduced-scale air model. PhD thesis, Department of Architecture, Massachusetts Institute of Technology, USA, p 211
68. Cropper P, Yang T, Cook M, Fiala D, Yousaf R (2010) Coupling a model of human thermoregulation with Computational Fluid Dynamics for predicting human-environment interaction. *J Building Perform Simulation* 3(3):233–243
69. Zhang H, Arens E, Huizenga C, Han T (2010) Thermal sensation and comfort models for non-uniform and transient environments, part III: whole-body sensation and comfort. *Building Environ* 45:399–410
70. Lomas KJ, Ji Y (2009) Resilience of naturally ventilated buildings to climate change: advanced natural ventilation and hospital wards. *Energy Buildings* 41(6):629–653
71. DoE (2010) Building energy software tools directory. Energy efficiency & renewable energy; Available from http://apps1.eere.energy.gov/buildings/tools_directory/. Accessed 12 June 2010
72. IBPSA-Germany (2010) Simupedia - A wiki about building simulation. Accessed 25 Aug 2009
73. Kolokotroni M (2007) Vent dis.course – Development of distance learning vocational training material for the promotion of best practice ventilation energy performance in buildings. European Commission Intelligent Energy – Europe Programme Publishable Final Report (EIE/04/022/S07.38630)
74. Healthheating (2010) Indoor environmental quality – Educational resource of the building and health sciences; Available from <http://www.healthyheating.com/index.htm>. Accessed 06 March 2010
75. Mumovic D, Santamouris M (eds) (2009) A handbook of sustainable building design & engineering. Earthscan, London
76. Pearson A (2011) Getting on the air . . . naturally. *CIBSE Journal*, Feb issue, pp 28–32. Available from <http://www.cibsejournal.com/issues/2011-02/>
77. Noble C (2011) Commerzbank: a sustainable skyscraper. Architecture 489; Available from http://web.utk.edu/~archinfo/a489_f02/PDF/commerzbank.pdf. Accessed 12 Feb 2011
78. Foster + Partners. Commerzbank Headquarters. Frankfurt, Germany; Available from <http://www.fosterandpartners.com/Projects/0626/Default.aspx>
79. BSI (2010) Constructing the business case – building information modelling. British Standard Institution and Building SMART UK, London
80. Alwaer H, Clements-Croome DJ (2010) Key performance indicators (KPIs) and priority setting in using the multi-attribute approach for assessing sustainable intelligent buildings. *Building Environ* 45(4):799–807

New Polymers, Renewables as Raw Materials

RICHARD P. WOOL

Department of Chemical Engineering, Center for Composite Materials, Affordable Composites from Renewable Sources (ACRES), University of Delaware, Newark, DE, USA

Article Outline

Glossary

Definition of the Subject

Introduction

Synthetic Pathways for Triglyceride-Based Monomers

Polymers from Plant Oils

Properties of Plant Oil Resins

Castor Oil-Based Polymer Properties

Future Directions

Bibliography

Glossary

Composite materials Are engineered or naturally occurring materials made from two or more constituent materials with significantly different physical or chemical properties which remain separate and distinct at the microscopic scale.

Lignin Is a complex chemical compound most commonly derived from wood and an integral part of the secondary cell walls of plants and some algae.

Triglyceride Is an ester derived from glycerol and three fatty acids. It is the main constituent of vegetable oil and animal fats.

Definition of the Subject

This entry describes how to make high volume materials out of thin-air using CO₂ and sunlight.

Introduction

Recent advances in genetic engineering, composite science, and natural fiber development offer significant opportunities for developing new, improved materials from renewable resources that can biodegrade or be recycled, enhancing global sustainability. A wide range

of high-performance, low-cost materials can be made using plant oils, natural fibers, and lignin. By selecting the fatty acid distribution function of plant oils via computer simulation and the molecular connectivity, chemical functionalization and molecular architecture can be controlled to produce linear, branched, or cross-linked polymers. These materials can be used as pressure-sensitive adhesives, elastomers, rubbers, foams, and composite resins. This entry describes the chemical pathways that were used to modify plant oils and allow them to react with each other and various comonomers to form materials with useful properties.

Polymers and polymeric composite materials have extensive applications in the aerospace, automotive, marine, infrastructure, military, sports, and industrial fields. These lightweight materials exhibit excellent mechanical properties, high corrosion resistance, dimensional stability, and low assembly costs. Traditionally, polymers and polymeric composites have been derived from petroleum; however, as the applications for polymeric materials increase, finding alternative sources of these materials has become critical. In recent years, the Affordable Composites from Renewable Sources (ACRES) program at the University of Delaware has developed a broad range of chemical routes to use natural triglyceride oils to make polymers and composite materials [1, 2]. These materials have economic and environmental advantages that make them attractive alternatives to petroleum-based materials.

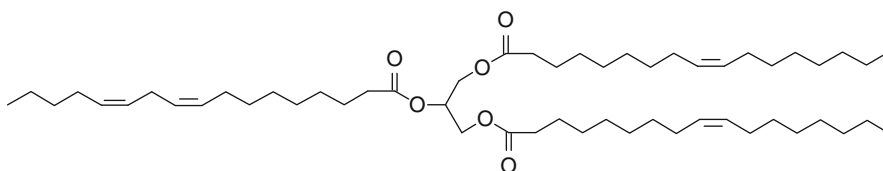
Natural oils, which can be derived from both plant and animal sources, are abundant in most parts of the world, making them an ideal alternative to chemical feedstocks. These oils are predominantly made up of triglyceride molecules, which have the structure shown in Fig. 1. Triglycerides are composed of three fatty acids joined at a glycerol juncture. The most common oils contain fatty acids that vary from 14 to 22 carbons in length with 0–3 double bonds per fatty acid. Table 1

shows the fatty acid distributions of several common oils [3]. Exotic oils are composed of fatty acids with other types of functionalities, such as epoxies, hydroxyls, cyclic groups, and furanoid groups [4]. Because of the many different fatty acids present, at a molecular level these oils are composed of many different types of triglycerides with numerous levels of unsaturation. With newly developed genetic engineering techniques, the variation in unsaturation can be controlled in plants such as soybean, flax, and corn; however, some oils are better suited to polymer resin development.

Besides applications in the foods industry, triglyceride oils have been used extensively to produce coatings, inks, plasticizers, lubricants, and agrochemicals [5–11]. In the polymers field, the use of these oils as toughening agents was investigated. Barrett [12] has reviewed an extensive amount of work on using these oils to produce interpenetrating networks (IPNs). It was found that IPNs formed by triglycerides could increase the toughness and fracture resistance in conventional thermoset polymers [13–16]; see also [17–25]. In these works, the functional triglyceride was a minor component in the polymer matrix acting solely as a modifier to improve the physical properties of the main matrix. Consequently, the triglyceride-based materials were low-molecular-weight (M_w), lightly cross-linked materials incapable of displaying the necessary rigidity and strength required for structural applications by themselves.

Synthetic Pathways for Triglyceride-Based Monomers

Triglycerides contain active sites amenable to chemical reaction: the double bond, the allylic carbons, the ester group, and the carbons alpha to the ester group. These active sites can be used to introduce polymerizable



New Polymers, Renewables as Raw Materials. Figure 1
Triglyceride molecule, the major component of natural oils

New Polymers, Renewables as Raw Materials. Table 1 Fatty acid distribution in various plant oils

Fatty acid	# C: # DB	Canola	Corn	Cottonseed	Linseed	Olive	Palm	Rapeseed	Soybean	High oleic ^a
Myristic	14:0	0.1	0.1	0.7	0.0	0.0	1.0	0.1	0.1	0.0
Myristoleic	14:1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Palmitic	16:0	1.1	10.9	21.6	5.5	13.7	41.4	3.0	11.0	6.4
Palmitoleic	16:1	0.3	0.2	0.6	0.0	1.2	0.2	0.2	0.1	0.1
Margaric	17:0	0.1	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0
Margaroleic	17:1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Stearic	18:0	1.8	2.0	2.6	3.5	2.5	1.1	1.0	1.0	3.1
Oleic	18:1	60.9	25.4	18.6	19.1	71.1	39.3	13.2	23.4	82.6
Linoleic	18:2	21.0	59.6	51.4	15.3	10.0	10.0	13.2	53.2	2.3
Linolenic	18:3	8.8	1.2	0.7	56.6	0.6	0.4	9.0	7.8	3.7
Arachidic	20:0	0.7	0.4	0.3	0.0	0.9	0.3	0.5	0.3	0.2
Gadoleic	20:1	1.0	0.0	0.0	0.0	0.0	0.0	9.0	0.0	0.4
Eicosadienoic	20:2	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0
Behenic	22:0	0.3	0.1	0.2	0.0	0.0	0.1	0.5	0.1	0.3
Erucic	22:1	0.7	0.0	0.0	0.0	0.0	0.0	49.2	0.0	0.1
Lignoceric	24:0	0.2	0.0	0.0	0.0	0.0	0.0	1.2	0.0	0.0
Average # DB/triglyceride		3.9	1.5	3.9	6.6	2.8	1.8	3.8	1.6	3.0

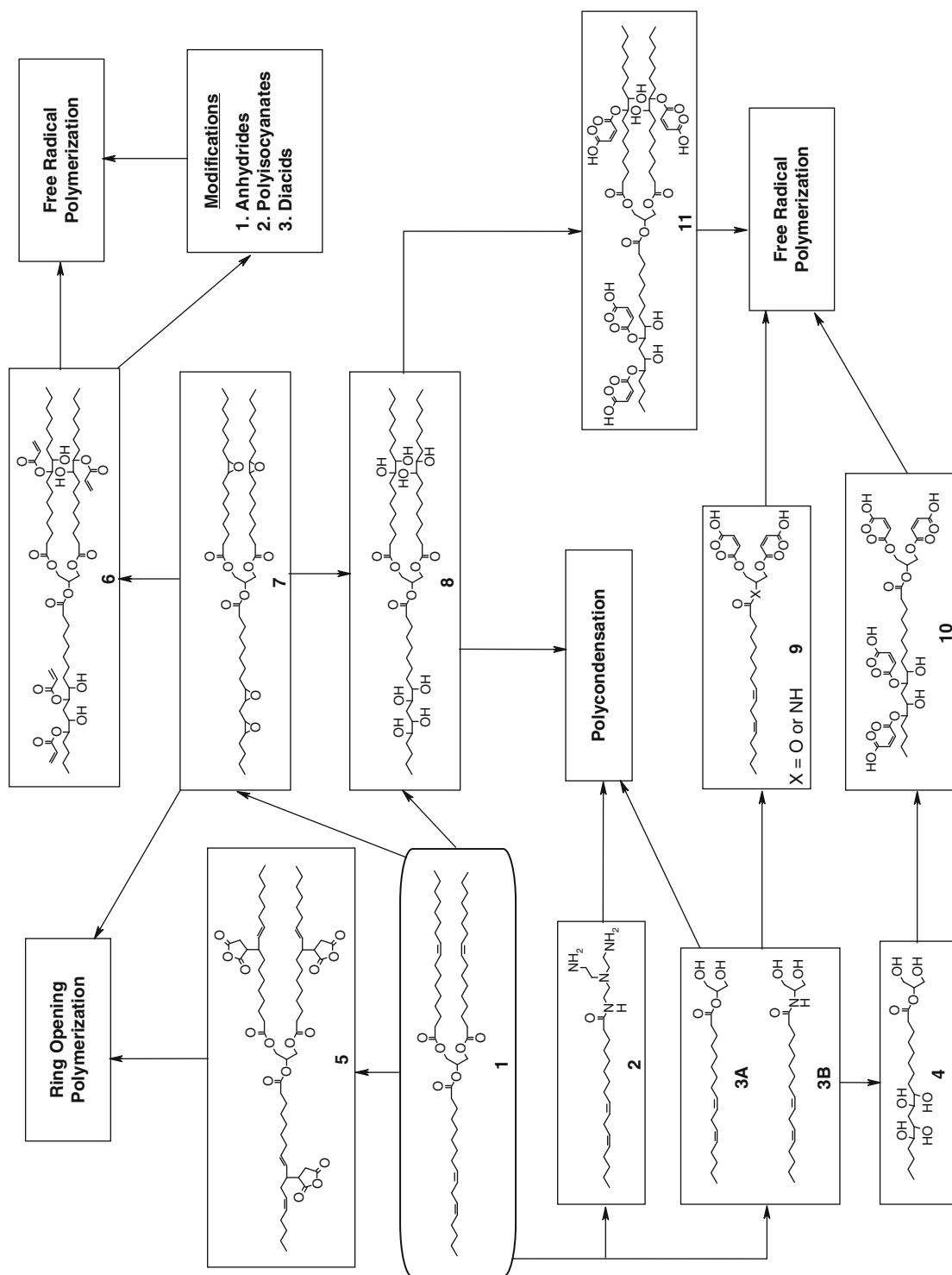
^aGenetically engineered high oleic acid content soybean oil (DuPont)

groups on the triglyceride using the same techniques applied in the synthesis of petrochemical-based polymers. The key step is to reach a higher level of M_w and cross-link density, as well as to incorporate chemical functionalities known to impart stiffness in a polymer network (e.g., aromatic or cyclic structures). Figure 2 illustrates several synthetic pathways that accomplish this [1].

In structures 5, 6, 7, 8, and 11, the double bonds of the triglyceride are used to functionalize the triglyceride with polymerizable chemical groups. From the natural triglyceride, it is possible to attach maleates (5) [6–11] or to convert the unsaturation to epoxy (7) [26–28] or hydroxyl functionalities (8) [29, 30]. Such transformations make the triglyceride capable of reaction via ring-opening or polycondensation polymerization. These particular chemical pathways are also accessible via natural epoxy and hydroxyl functional triglycerides [12, 14–16]. It is also possible to attach

vinyl functionalities to the epoxy and hydroxyl functional triglycerides. Reaction of the epoxy functional triglyceride with acrylic acid incorporates acrylates onto the triglyceride (6), while reaction of the hydroxylated triglyceride with maleic anhydride incorporates maleate half-esters and esters onto the triglyceride (11). These monomers can then be blended with a reactive diluent, similar to most conventional vinyl ester resins and cured by free-radical polymerization.

The second method for synthesizing monomers from triglycerides is to convert the triglyceride to monoglycerides through a glycerolysis (3A) reaction or an amidation reaction (2, 3B) [31–36]. Monoglycerides are used in surface coatings, commonly referred to as alkyd resins, because of their low cost and versatility [32]. In those applications, the double bonds of the monoglyceride are reacted to form the coating. However, monoglycerides also can react through the alcohol groups via polycondensation reactions with



New Polymers, Renewables as Raw Materials. Figure 2
Chemical pathways leading to polymers from triglyceride molecules [1]

a comonomer, such as a diacid, epoxy, or anhydride. Alternatively, maleate half-esters can be attached to these monoglycerides (**9**) allowing them to undergo free-radical polymerization.

The third method is to functionalize the unsaturation sites as well as reduce the triglyceride into monoglycerides. This can be accomplished by glycerolysis of an unsaturated triglyceride, followed by hydroxylation or by glycerolysis of a hydroxy functional triglyceride. The resulting monomer can then be reacted with maleic anhydride, forming a monomer capable of polymerization by the free-radical mechanism [1].

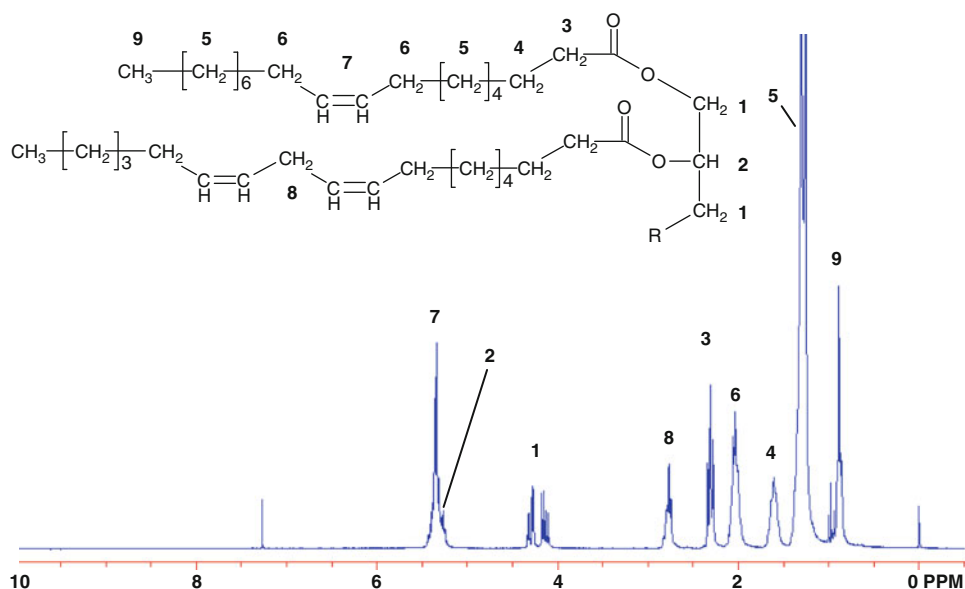
Although the structure of triglycerides is complex in nature, it is possible to characterize some aspects of it using proton nuclear magnetic spectroscopy (^1H -NMR) and Fourier transform infrared (FTIR) spectroscopy. A typical ^1H NMR spectrum of soybean oil is shown in Fig. 3, with peak assignments. The two sets of peaks at 1.0–1.4 ppm are produced by the four glycerol methylene protons per triglyceride [0]. The triplet set of peaks at 2.3 ppm is produced by the six protons in the alpha position relative to the carbonyl groups. The peak at 0.9 ppm is produced by the nine methyl protons per triglyceride at the end of each fatty acid chain. These three groups of peaks provide

a standard by which other peaks can be used to quantitatively characterize functional groups in the triglyceride.

In this entry, the focus is on three triglyceride monomers, shown in Fig. 4, which have been found to be promising candidates for use in the composites and engineering plastics fields. They are acrylated epoxidized soybean oil (AESO), the maleinized soybean oil monoglyceride (SOMG/MA), and maleinized hydroxylated soybean oil (HSO/MA). These monomers, when used as a major component of a molding resin, have shown properties comparable to conventional polymers and composites, and these properties will be presented. In addition, they can be used as a matrix in synthetic and natural fiber-reinforced composites.

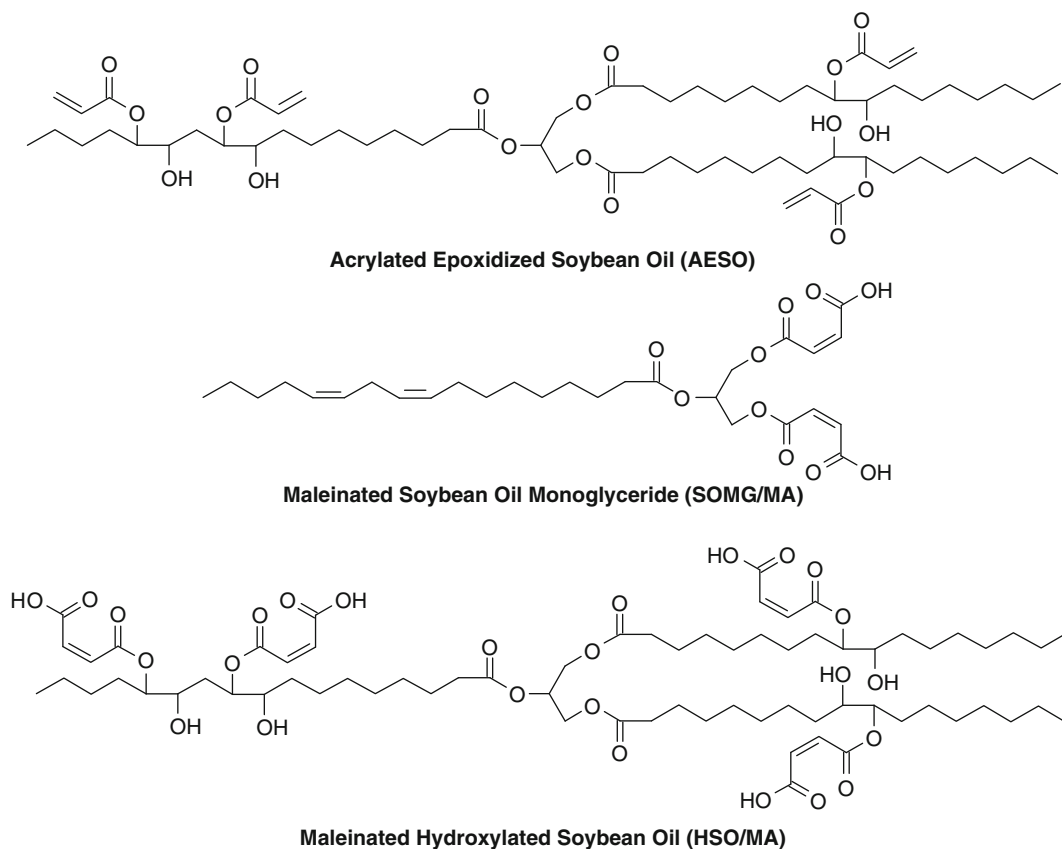
Acrylated Epoxidized Soybean Oil

Acrylated epoxidized oils (Fig. 4) are synthesized from the reaction of acrylic acid with epoxidized triglycerides. Epoxidized triglycerides can be found in natural oils, such as vernonia plant oil, or can be synthesized from more common unsaturated oils, such as soybean oil or linseed oil, by a standard epoxidation reaction [37]. The natural epoxy oil, vernonia oil, has



New Polymers, Renewables as Raw Materials. Figure 3

^1H -NMR spectrum of soybean oil (CDCl_3). R represents a third fatty acid



New Polymers, Renewables as Raw Materials. Figure 4
Triglyceride based monomers

a functionality of 2.8 epoxy rings per triglyceride [4]. Epoxidized soybean oil is commercially available and is generally sold with a functionality of 1.1–1.6 epoxy rings per triglyceride, which can be identified via ^1H NMR [20, 38]. Epoxidized linseed oil is also commercially available when higher epoxy content is required. Predominantly, these oils are used as alternative plasticizers in polyvinyl chloride in place of phthalates [39–41], but their use as a toughening agent also was explored [20, 23–25, 42]. With the addition of acrylates, the triglyceride can be reacted via addition polymerization. AESO was used extensively in surface coatings and is commercially manufactured in forms such as Ebecryl 860 [7, 43, 44]. Urethane and amine derivatives of AESO have also been developed for coating and ink applications [8, 9, 45].

The reaction of acrylic acid with epoxidized soybean oil occurs through a standard substitution

reaction and was found to have first-order dependence with respect to epoxy concentration and second-order dependence with respect to acrylic acid concentration [46]. However, epoxidized oleic methyl ester was found to display second-order dependence on both epoxy and acrylic acid concentrations [47]. Although the reaction of epoxidized soybean oil with acrylic acid is partially catalyzed by the acrylic acid, the use of additional catalysts is common. Tertiary amines, such as *N,N*-dimethylaniline, triethylamine, and 1,4-diazobicyclo[2.2.2]octane, are commonly used [38, 48]. Additionally, more selective organometallic catalysts have been developed that reduce the amount of epoxy homopolymerization [49, 50].

AESO can be blended with a reactive diluent, such as styrene, to improve its processability and to control the polymer properties to reach a range acceptable for structural applications. By varying the amount of

styrene, it is possible to produce polymers with different moduli and glass transition temperatures. Polymer properties can also be controlled by changing the M_w of the monomer or the functionality of the acrylated triglyceride. Consequently, a range of properties, and therefore applications, can be found. Subsequent to the acrylation reaction, the triglyceride contains both residual amounts of unreacted epoxy rings as well as newly formed hydroxyl groups, both of which can be used to further modify the triglyceride by reaction with a number of chemical species, such as diacids, diamines, anhydrides, and isocyanates. The approach presented here is to oligomerize the triglycerides with reagents that have chemical structures conducive to stiffening the polymer, such as cyclic or aromatic groups. Reacting AESO with cyclohexanedicarboxylic anhydride (Fig. 5a) forms oligomers, increasing the entanglement density as well as introducing stiff cyclic rings to the structure. Reaction of the AESO with maleic acid (Fig. 5b) also forms oligomers and introduces more double bonds. Although it is desirable to maximize the conversion of hydroxyls or epoxies, the viscosity increases dramatically at high levels of conversion. Eventually, this can lead to gelation, so the reaction must be carefully monitored. After oligomerization, the modified AESO resin can be blended with styrene and cured in the same manner as the unmodified AESO resin.

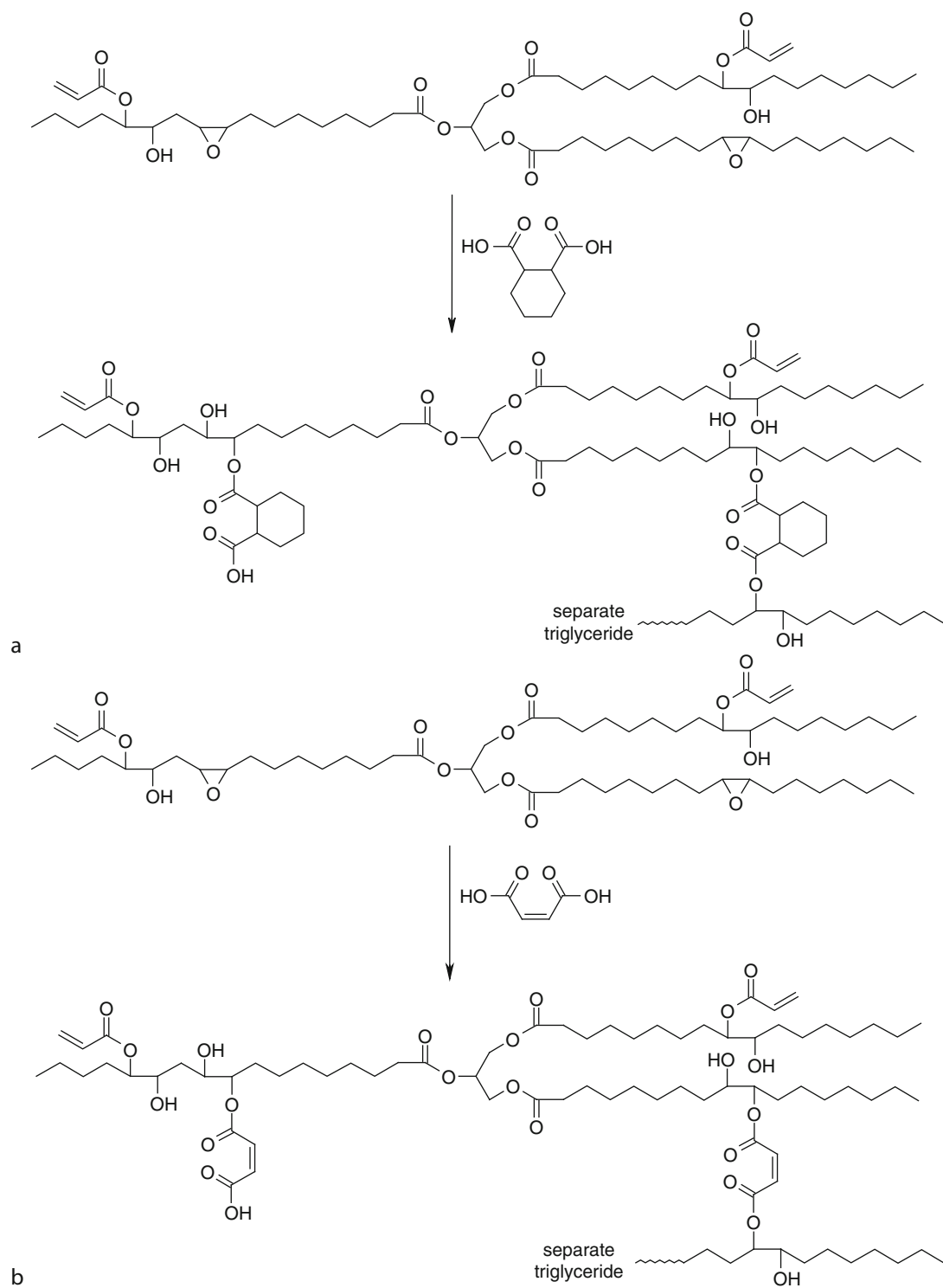
Maleinized Soybean Oil Monoglyceride

Maleinized soybean oil monoglyceride (Fig. 4) is synthesized from the triglyceride oil in two steps [33]. The first is a standard glycerolysis reaction to convert the triglycerides into monoglycerides by reacting triglycerides with glycerol; see [31]. The product is generally a mixture of mono- and diglycerides, as illustrated in Fig. 6. Using excess glycerol can aid in conversion. Additionally, the reaction can be run in solvent or in the presence of an emulsifier catalyst [34]. Once the reaction is completed, it is possible to separate a portion of the unreacted glycerol by cooling the product rapidly [33]. The presence of glycerol is not detrimental to the end polymer, because it can be reacted with maleic anhydride in the same manner as the monoglycerides and incorporated into the end polymer network.

The maleinization of the soybean oil monoglyceride (SOMG) mixture at temperatures below 100°C produces monoglyceride, diglyceride, and glycerol maleate half-esters. This reaction makes no attempt to produce a polyester, and the half-ester formation is expected to proceed at low temperatures in the presence of either acid or base catalysts without any by-products. A good indication of the success of this reaction is to follow the signal intensity ratio of maleate vinyl protons to fatty acid vinyl protons (N_M/N_{FA}) in the 1H NMR spectrum. The use of 2-methylimidazole and triphenyl antimony as catalysts was shown to be successful when conducting the reaction at temperatures of 80–100°C with a 3:2 weight ratio of glycerides to maleic anhydride ($N_M/N_{FA} = 0.85$) [33, 51]. Once these maleates have been added, the monoglycerides can react via addition polymerization. Because maleates are relatively unreactive with each other, the addition of styrene increases the polymerization conversion and imparts rigidity to the matrix.

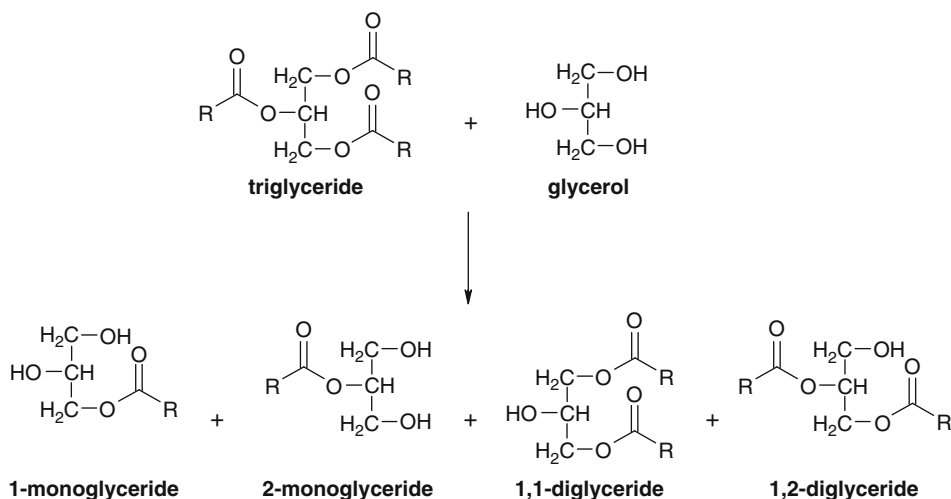
To increase the glass transition temperature (T_g) and modulus of the SOMG/MA polymer, more rigid diols can be added during the maleinization reaction. Two such diols are neopentyl glycol (NPG) and bisphenol A (BPA), which may increase the rigidity of the end polymer network. Although their addition to the maleinization mixture will reduce the renewable resource content of the final resin, they should result in higher T_g values for the end polymer. The synthesis of maleate half-esters of organic polyols, including NPG and BPA and the cross-linking of the resulting maleate half-esters with a vinyl monomer such as styrene, has been reported [52, 53]. The literature also abounds with examples of unsaturated polyesters prepared from NPG and maleic anhydride with other polyols and diacids [54–57]. However, the copolymers of NPG and BPA bis maleate half-esters with SOMG maleate half-esters are new.

The properties of the SOMG/MA polymer as well as the effect of adding NPG and BPA on the mechanical properties of the final polymers are presented here. For this purpose, mixtures of SOMG/NPG and SOMG/BPA, prepared at the same weight ratio, were maleinized, and the copolymers of the resulting maleates with styrene were analyzed for their mechanical properties and compared to that of SOMG maleates.



New Polymers, Renewables as Raw Materials. Figure 5

(a) Modification of AESO by reaction with cyclohexane dicarboxylic anhydride. **(b)** Modification of AESO by reaction with maleic acid



New Polymers, Renewables as Raw Materials. Figure 6
Glycerolysis of triglycerides to form mixtures of mono- and diglycerides

Maleinized Hydroxylated Oil

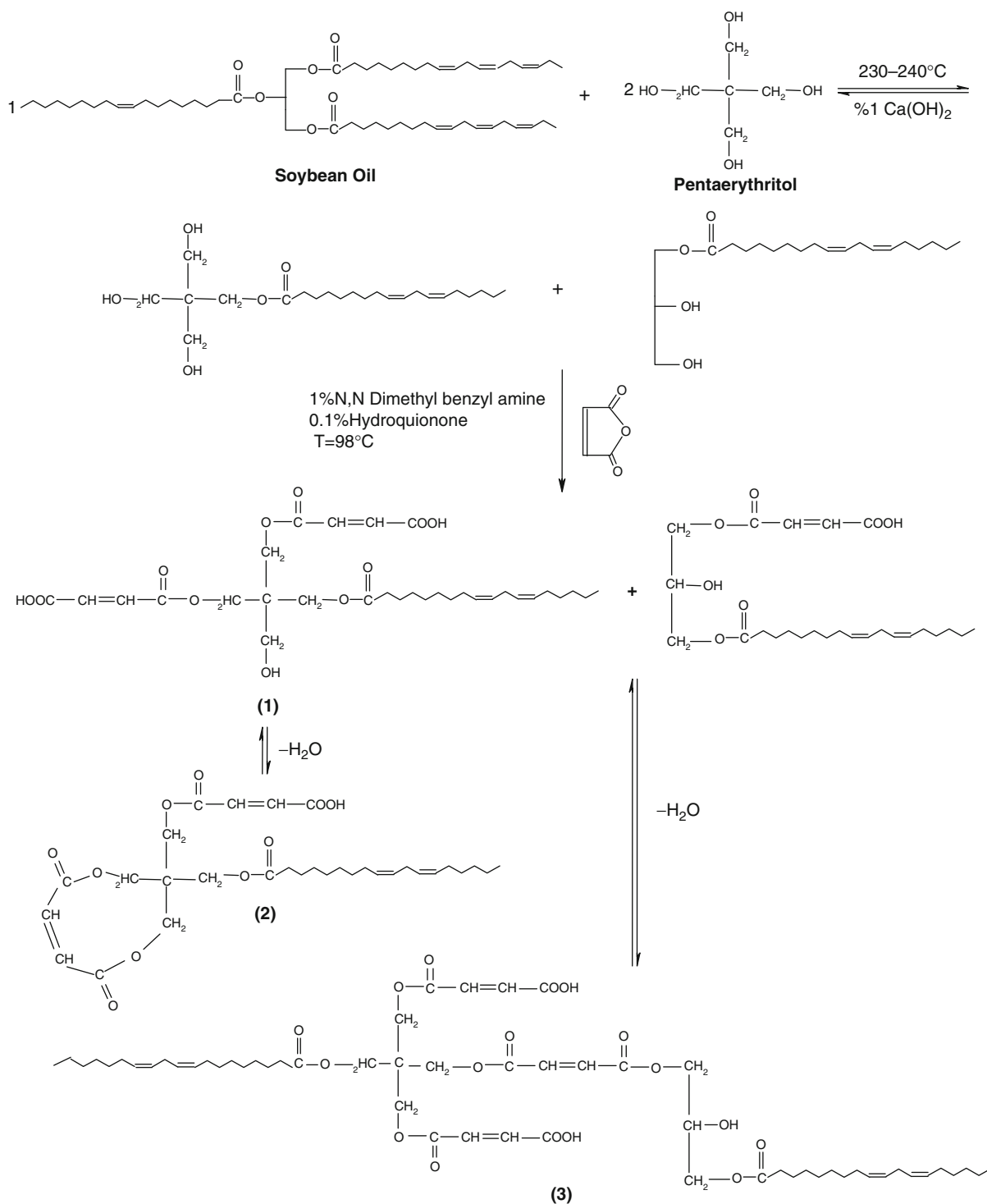
Maleinized hydroxylated oil (HO/MA) is synthesized in a manner similar to both the AESO and the SOMG/MA monomers. The double bonds of an unsaturated oil are used to attach the polymerizable groups by converting the double bonds of the triglyceride to hydroxyl groups. The hydroxyls can then be used to attach maleates. As shown in Fig. 2, there are two routes to synthesize the hydroxylated triglyceride. The first is through an epoxidized intermediate. By reacting the epoxidized triglyceride with an acid, the epoxies can be easily converted to hydroxyl groups [29, 58]. Alternatively, the hydroxylated oil can be synthesized directly from the unsaturated oil, as described in [1]. After hydroxylation, the oil can be reacted with maleic anhydride to functionalize the triglyceride with maleate half-esters. A molar ratio of 4:1 anhydride to triglyceride was used in all cases, and the reaction was catalyzed with *N,N*-dimethylbenzylamine. Once the maleinization reaction is finished, the monomer resin can be blended with styrene similar to the other resins presented here.

SOPERMA: Soybean Oil Pentaerythritol Glyceride Maleates

In the preceding sections, the preparation of soybean oil monoglyceride maleates (SOMG/MA) by E. Can [65] was reported. Soybean oil was reduced to

monoglycerides through a glycerolysis reaction, and the glycerolysis product, which was a mixture of mono- and diglycerides as well as unreacted triglycerides and free glycerol, was reacted with maleic anhydride to convert the free hydroxyls to maleate half-esters, thus allowing them to free radically polymerize. The use of pentaerythritol instead of glycerol in the same synthetic route, as shown in Fig. 7, offers certain advantages, such as the presence of more hydroxyl groups and, therefore, more reactive sites for malination, which should result in a higher cross-link density for the resulting polymers. There are no standard conditions for the alcoholysis of the triglyceride oils with pentaerythritol; the reactant molar ratios and reaction conditions change according to the end use of the product. Soybean pentaerythritol molar ratios of SO:PER = 1:2 and SO:PER = 1:3 were used, which should give mixtures of monoglycerides and pentaerythritol monoesters as the main products. Higher amounts of pentaerythritol were avoided because this would further decrease the triglyceride content of the formulation. The idealized reaction scheme for the soybean oil pentaerythritol alcoholysis reaction is shown in Fig. 7.

The soybean oil pentaerythritol alcoholysis reactions were carried out for the first time by E. Can at 230–240°C for 0.5, 2, and 5.5 h. A reaction temperature of 180–190°C was also employed. Ca(OH)₂ was used as



New Polymers, Renewables as Raw Materials. Figure 7

The reaction scheme for soybean oil pentaerythritol alcoholysis and malination reaction (SOPERMA)

a catalyst at a concentration of 1% of the total weight of the oil and the polyol. $\text{Ca}(\text{OH})_2$ forms soap with the free fatty acids in the oil and promotes the reaction at least in part by increasing the solubility of pentaerythritol in the oil. $\text{Ca}(\text{OH})_2$ was reported to be an effective catalyst for the glycerolysis of soybean oil; it increases the monoglyceride yield and reduces the triglyceride content of the glycerolysis product. The amount of reactants and catalysts used for the soybean oil pentaerythritol alcoholysis reactions and the resulting malination reactions, at different mole ratios, are shown in Table 2.

For the malination of the alcoholysis products, maleic anhydride was in a 1:1 molar ratio with the number of hydroxyls on pentaerythritol used in the alcoholysis reaction, thus molar ratios of SO:PER:MA = 1:2:8 and 1:3:12 were employed. A molar ratio of SO:PER:MA = 1:3:7.62 was also used to reduce the unreacted maleic anhydride content of the latter reaction. A reaction temperature of 95–100°C was used, because when lower reaction temperatures were used, for example, 60°C, the reaction rate was significantly lower and therefore not preferred. Temperatures above 100°C led to gelation of the product due to polyesterification of both the maleate half-esters and maleic anhydride with the free hydroxyls and were therefore avoided. Table 2 shows the amount of the reactants used for the malination reactions done at different SO:PER:MA molar ratios. The product at room temperature was a light brown solid. The SOPERMA (1:3:12) and SOPERMA (1:2:8) products were prepared

in a similar manner by changing the molar ratios of the reactants in the formulation.

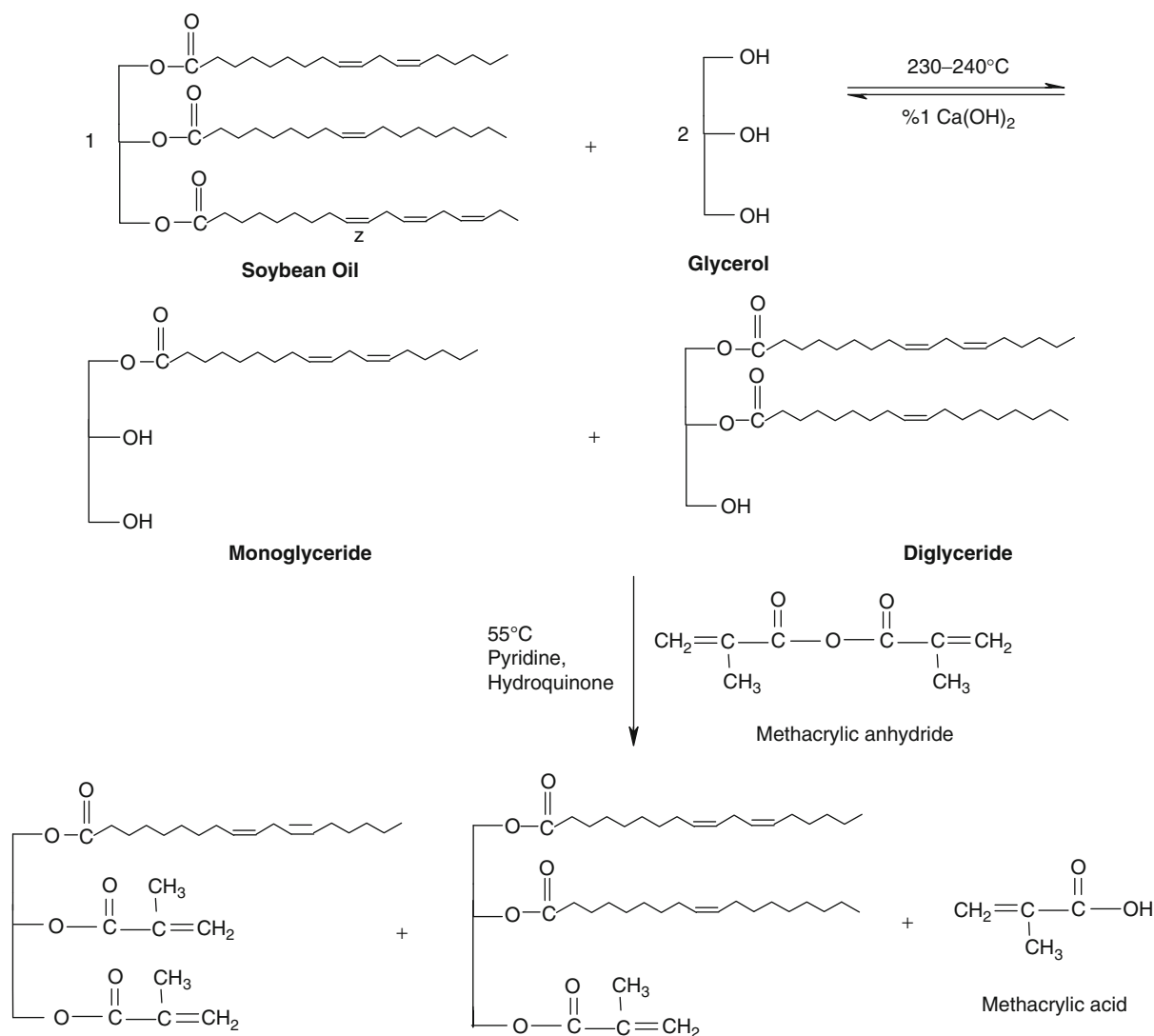
SOGLYME: Soybean Oil Monoglyceride Methacrylates

Soybean oil monoglyceride methacrylates (SOGLYME) were synthesized by E. Can [66]. Soybean oil monoglyceride methacrylates were prepared in a two-step process. First, soybean oil was glycerolized in the presence of $\text{Ca}(\text{OH})_2$ as catalyst at 230–240°C for 5 h. The glycerolysis of soybean oil under these conditions gives a product with an equilibrium mixture containing the monoglycerides, diglycerides, and the two starting materials [63]. The glycerolysis product was then reacted with methacrylic anhydride at 55°C to form the methacrylate esters of the glycerides and methacrylic acid. Pyridine, which is an effective catalyst in the reaction of methacrylic anhydride with alcohols, was used as the catalyst. Hydroquinone was used to inhibit the radical polymerization of the reactive methacrylate esters. The idealized reaction schemes for the glycerolysis and methacrylation are shown in Fig. 8.

The glycerolysis of soybean oil was carried out in a 1:2.4 molar ratio of soybean oil to glycerol. For the following methacrylation reaction, methacrylic anhydride was in a 1:1 molar ratio with the number of hydroxyls on glycerol used in the glycerolysis reaction, thus the molar ratio of soybean oil, glycerol, and methacrylic anhydride was SO:GLYC:ME = 1:2.4:7.2. The amount of reactants and catalysts used for the

New Polymers, Renewables as Raw Materials. Table 2 The amount of ingredients used in the alcoholysis of soybean oil with pentaerythritol and the following malination reactions

Reactants	SO:PER:MA (1:2:8)		SO:PER:MA (1:3:12)		SO:PER:MA (1:3:7.62)	
	Weight of reactant (g)	Moles of reactant	Weight of reactant (g)	Moles of reactant	Weight of reactant (g)	Moles of reactant
Soybean oil	5	0.0057	5	0.0057	5	0.0057
Pentaerythritol	1.554	0.01143	2.331	0.01714	2.331	0.01714
$\text{Ca}(\text{OH})_2$	0.0655	–	0.0733	–	0.0733	–
Maleic anhydride	1.479	0.04571	6.720	0.06857	1.267	0.0435
N,N dimethyl benzyl amine	0.1103	–	0.1405	–	0.116	–
Hydroquinone	0.011	–	0.014	–	0.0116	–



New Polymers, Renewables as Raw Materials. Figure 8

The reaction scheme for soybean oil glycerolysis and methacrylation reactions (SOGLYCMA)

soybean oil glycerolysis and methacrylation reactions are shown in Table 3. The resulting SOGLYME product was a light yellow liquid.

COPERMA: Castor Oil Pentaerythritol Glyceride Maleates

Castor oil is not commonly used in alkyd resin formulations and there are few reports on the alcoholysis of castor oil triglycerides. For the preparation of castor oil-based monomers, castor oil was first alcoholized with glycerol, pentaerythritol, and an aromatic diol;

BPA propoxylate and the alcoholysis products were then malinated, as shown in Fig. 9 [66]. Bisphenol A propoxylate was used specifically to introduce the rigid aromatic rings onto the triglyceride structure. The maleate esters of castor oil alcoholysis products have never been synthesized before, thus the castor oil-based monomers presented here are totally new resins. The alcoholysis reactions of castor oil were carried out for 2 h at 230–240°C in the presence of Ca(OH)₂ as catalyst, similar to the soybean oil alcoholysis reactions. The malination reactions were carried out for 5 h at

New Polymers, Renewables as Raw Materials. Table 3
Reactant amounts used in soybean oil glycerolysis and methacrylation reactions (SO:GLYC:ME = 1:2.4:7.2)

Component	Weight(g)	Moles
Soybean oil	119.93	0.1371
Glycerol	30.30	0.329
Calcium hydroxide	0.7514	–
Methacrylic anhydride	152.14	0.987
Pyridine	3.024	–
Hydroquinone	0.302	–

98°C to ensure the completeness of the malination of the secondary hydroxyls of castor oil. *N,N*-Dimethylbenzylamine, which is reported to be an effective catalyst for the malination of hydroxylated oils, was used as a catalyst. Castor oil was also directly malinated to see the effect of the alcoholysis step on the mechanical properties of the resulting polymers.

The molar ratio of castor oil to maleic anhydride was 1:3 for malination of castor oil; therefore, the reaction was carried out in an excess of maleic anhydride assuming that 1 mol castor oil contains 2.7 mol of hydroxyls. The reactants used in this reaction as well as their mole numbers and masses are given in Table 1. The COPERMA product was a light brown solid (Table 4).

COGLYMA: Castor Oil Monoglyceride Maleates

The castor oil glycerolysis reaction was carried out in a molar ratio of CO:GLY = 1:2.2 using reactants shown in Table 5. The reaction mixture was heated to 230–240°C and agitated under N₂ atmosphere for 2 h at this temperature. The reaction product at room temperature was a light brown liquid. The idealized structures of both the reactants and products for the glycerolysis reaction are shown in Fig. 10. Both unreacted castor oil and excess glycerol exist as by-products in the reaction.

For the malination of the castor oil glycerolysis product, maleic anhydride was used in a molar ratio sufficient to malinate the hydroxyls of both castor oil and the glycerol used in the glycerolysis reaction. Thus, the molar ratio of castor oil (CO), glycerol (GLY), and

maleic anhydride (MA) was CO:GLY:MA = 1:2.2:9.3. The castor oil glycerolysis product as prepared above was heated to 90°C with mechanical stirring, then the specified amounts of maleic anhydride and hydroquinone were added. The mixture was stirred at this temperature until the maleic anhydride melted and mixed with the castor oil glycerolysis product. *N,N*-Dimethylbenzylamine was added and the reaction mixture was heated to 98°C. The mixture was agitated at this temperature for 5 h. The reaction product at room temperature was a light brown solid.

COBPPRMA: Castor Oil Bisphenol a Propoxylate Glyceride Maleates

The castor oil (CO), pentaerythritol (PER) alcoholysis reaction was carried out in a molar ratio of CO:PER = 1:2 [66]. The reaction product at room temperature was a light brown liquid. The idealized structures of both the reactants and products for the alcoholysis reaction are shown in Fig. 11. Both unreacted castor oil and excess glycerol exist as by-products in the reaction. For the malination of the castor oil pentaerythritol alcoholysis product, maleic anhydride was used in a molar ratio to malinate the hydroxyls of both castor oil and the pentaerythritol used in the alcoholysis reaction. Thus, the molar ratio of castor oil (CO), pentaerythritol (PER), and maleic anhydride (MA) was CO:PER:MA = 1:2:10.7. The malinated product at room temperature was a light brown solid.

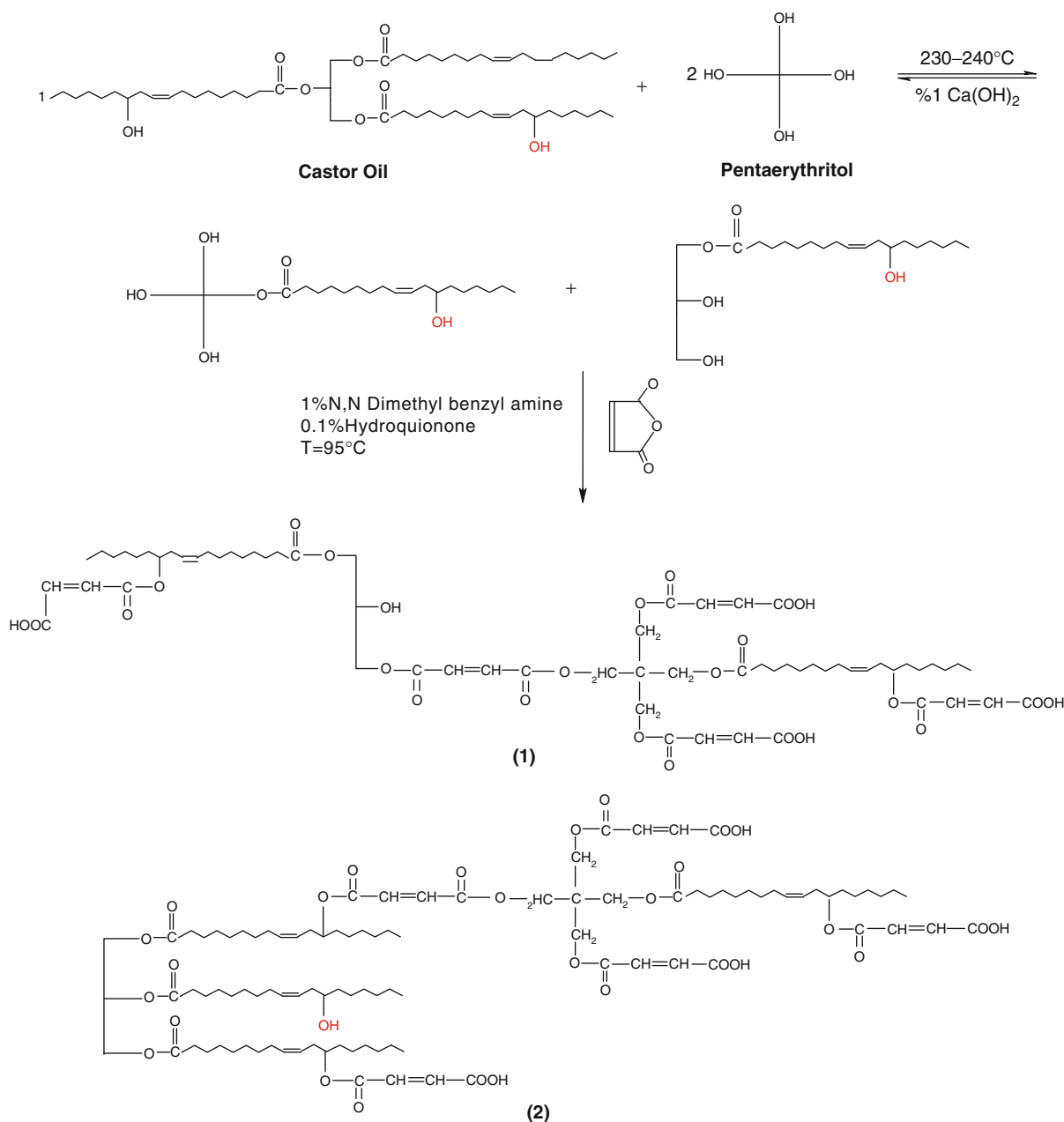
COMA: Castor Oil Maleates

COMA consists of the castor oil (CO) which was directly maleated on the three hydroxyl groups using maleic anhydride (MA). The ratios of reactants used are shown in Table 6 [66].

Polymers from Plant Oils

Acrylated Epoxidized Soybean Oil Polymers

Acrylated epoxidized soybean oil (AESO), shown in Fig. 4, was examined for its ability to produce high *T_g* and high modulus polymers. A commercial form of AESO, Ebecryl 860, was blended with various amounts of styrene to determine the effect of blending on mechanical and dynamic mechanical properties. The AESO used had an average functionality of



New Polymers, Renewables as Raw Materials. Figure 9

The reaction scheme for castor oil pentaerythritol alcoholysis and malination reactions (COPERMA)

approximately three acrylates per triglyceride as determined by ^1H NMR [38]. The optimal number of acrylates per triglyceride to obtain maximum stiffness and strength is about five acrylic acid groups per triglyceride. An example ^1H NMR spectrum of AESO is shown in Fig. 12. Similar to soybean oil, the triplet peak at

2.3 ppm can be used as a basis for the protons present alpha to the carbonyls in the triglyceride. The three peaks in the range of 5.8–6.5 ppm represent the three protons of the acrylate group.

Styrene monomer was blended with the AESO along with a free-radical initiator, 2,5-dimethyl-2,5-di

New Polymers, Renewables as Raw Materials. Table 4
Reactant amounts used in castor oil pentaerythritol alcoholysis and malination reactions (CO:PER:MA = 1:2:10.7)

Component	Weight(g)	Moles
Castor oil	120	0.13
Pentaerythritol	35.33	0.26
Calcium hydroxide	0.778	
Maleic anhydride	178.99	1.826
N,N dimethyl benzyl amine	3.35	
Hydroquinone	0.335	

New Polymers, Renewables as Raw Materials. Table 5
Reactant amounts used in castor oil glycerolysis and malination reactions (CO:GLY:MA = 1:2.2:9.3)

Component	Weight(g)	Moles
Castor oil	100	0.108
Glycerol	21.9	0.238
Calcium hydroxide	1.22	–
Maleic anhydride	98.08	1.001
N,N dimethyl benzyl amine	2.20	–
Hydroquinone	0.220	–

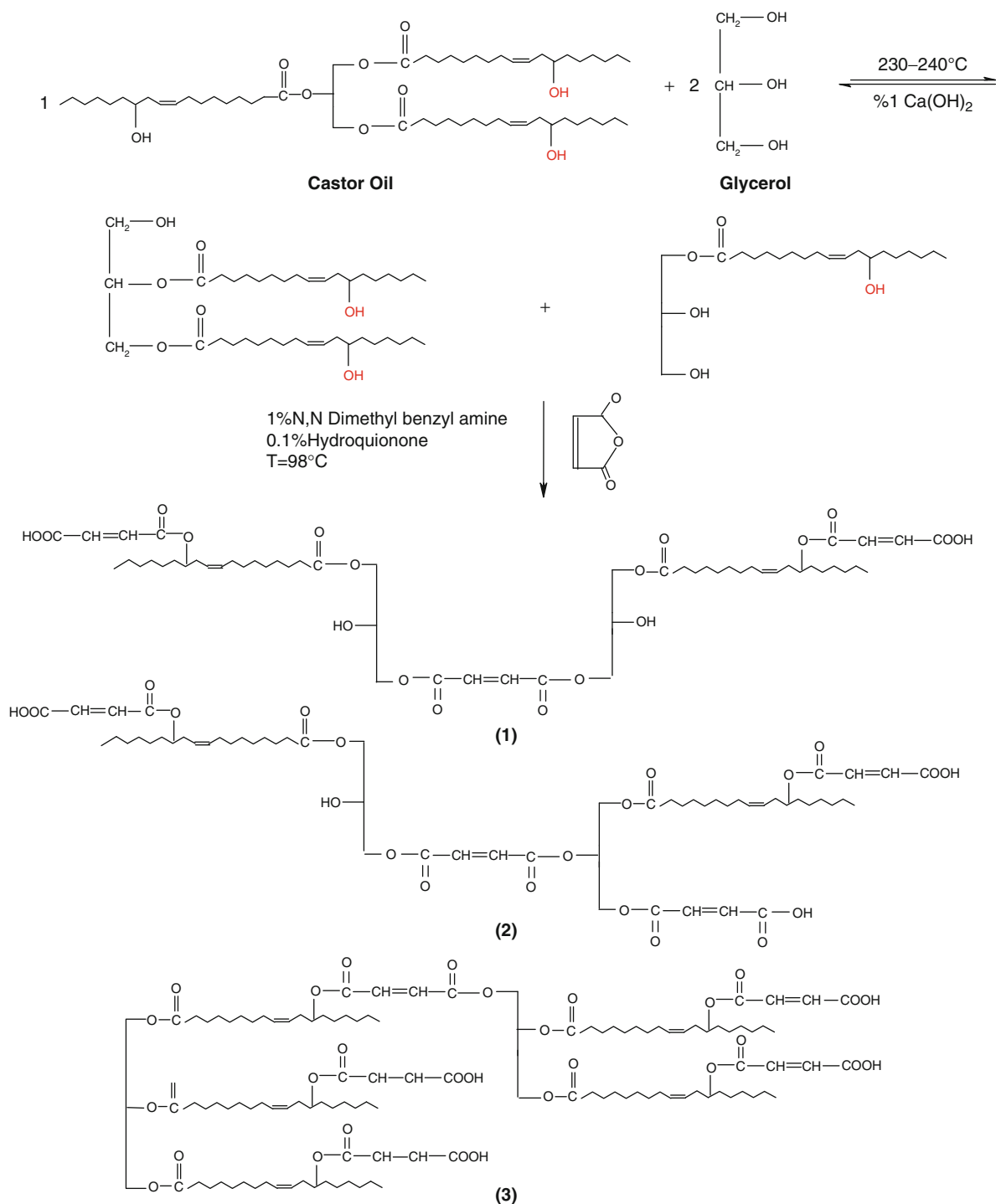
(2-ethylhexanoylperoxy) hexane. The addition of styrene to any type of unsaturated polyester is common practice in the composite liquid molding resin field. Its low cost and low viscosity improve the price and processability of the resin. For triglyceride-based polymers, the styrene also imparts a rigidity that the triglyceride does not naturally possess. The amount of initiator used was 1.5 wt% of the total resin weight (AESO plus styrene). For tensile testing of the polymers, samples were prepared in accordance with ASTM D 638. The resin was cured at 60°C for 12 h, followed by 125°C for 1.5 h. Samples for dynamic mechanical analysis (DMA) testing were prepared by pouring resin into a rubber gasket between two metal plates covered with aluminum foil. Samples were cured at 65°C for 1.5 h and postcured at 125°C for 1.5 h.

Synthesis of Modified Acrylated Epoxidized Soybean Oil Polymers

To improve the properties of the AESO-based resins, modified forms of the AESO were synthesized. These modifications involved partially reacting epoxidized soybean oil with acrylic acid and reacting the remaining epoxies with anhydrides or diacids. A more detailed explanation of the synthesis of partially acrylated epoxidized soybean oil can be found in other sources [38]. In summary, a mixture of epoxidized soybean oil was mixed with a stoichiometric amount of acrylic acid (about 1,500 g ESO to 460 g acrylic acid). Hydroquinone was added as a free-radical inhibitor in the amount of 0.07 wt% of the total reactants' weight, as well as 1,4-diazobicyclo[2.2.2]octane to act as a catalyst in the amount of 0.1 wt% of the total reactants' weight. This was reacted at 95°C for about 11 h, after which it was allowed to cool to room temperature. The resulting product had approximately 1.7 acrylates/triglyceride and 0.4 residual epoxy/triglyceride according to ¹H NMR. The remaining 2.3 epoxies were lost to epoxy homopolymerization [38].

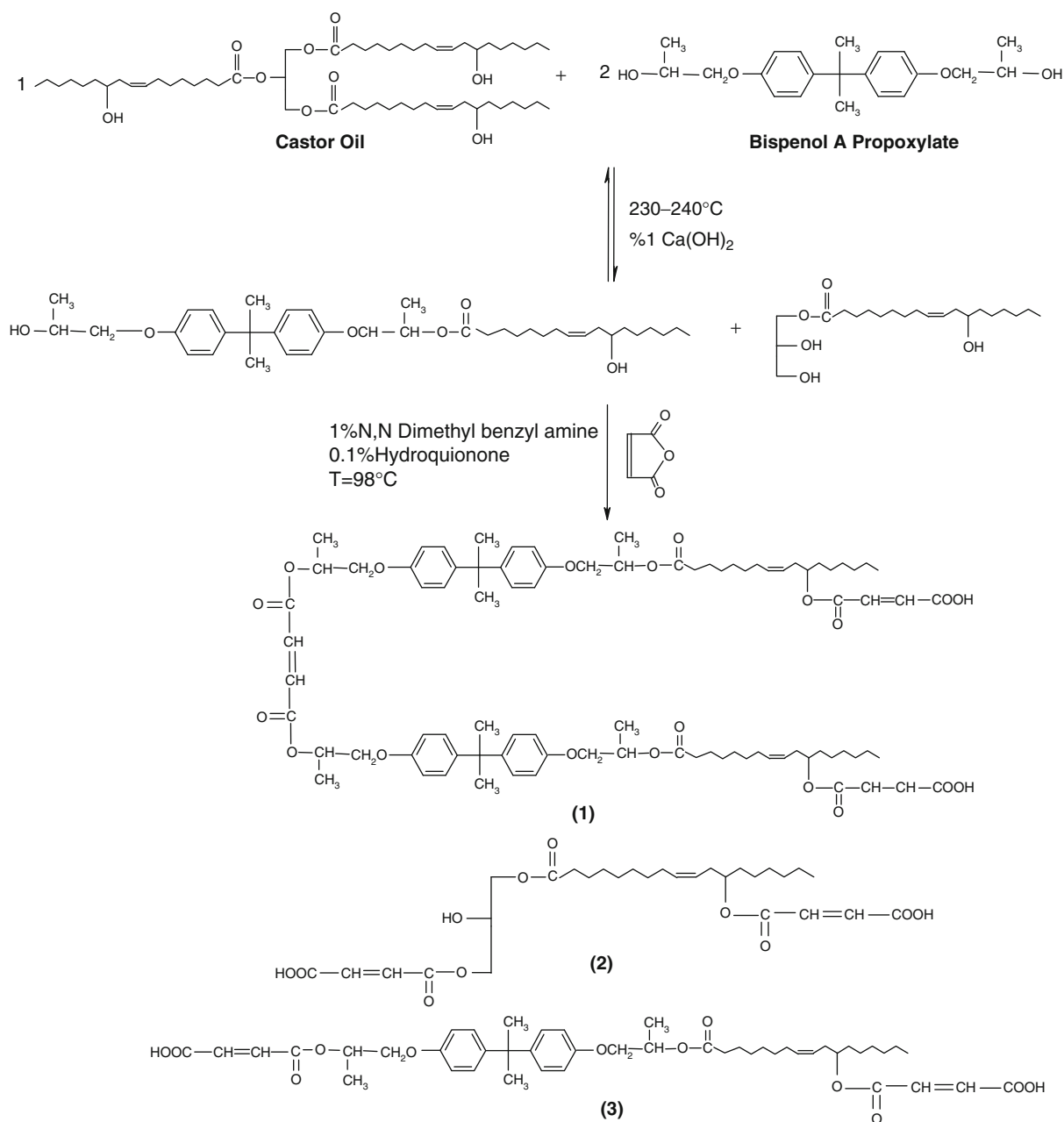
The first modification was the reaction of AESO with cyclohexanedicarboxylic anhydride (CDCA), as illustrated earlier in Fig. 5a. In a typical reaction, the synthesized AESO was reacted with 7.4% of its weight in CDCA and 0.1% of its weight in 2-methyl imidazole, which catalyzes the reaction [38]. After reacting at 110°C for about 3 h, the majority of the anhydride and epoxy groups was consumed, as indicated by FTIR spectroscopy. The second modification was the reaction of AESO with maleic acid (Fig. 5b). This was accomplished by reacting the synthesized AESO with 11% of its weight in maleic acid [38]. The reaction was held at approximately 80°C for 4 h, during which consumption of the epoxies was again confirmed by FTIR spectroscopy.

The modified resins were then blended with styrene and initiator in the amounts of 66 wt% modified AESO, 33 wt% styrene, and 1 wt% 2,5-dimethyl-2,5-di(2-ethylhexanoylperoxy) hexane initiator. After curing at 65°C for 1.5 h and postcuring at 125°C for 1.5 h, the polymers' dynamic mechanical properties were analyzed and compared to the unmodified AESO resin.



New Polymers, Renewables as Raw Materials. Figure 10

The reaction scheme for castor oil glycerolysis and malination reactions (COGLYMA)



New Polymers, Renewables as Raw Materials. Figure 11

The reaction scheme for castor oil Bisphenol A propoxylate alcoholysis and malination reactions (COBPAPRMA)

Maleinized Soybean Oil Monoglyceride Resin Synthesis

The maleinized soybean oil monoglyceride (Fig. 4) was synthesized by breaking the triglycerides into monoglyceride and then functionalizing the alcohol groups

with maleic anhydride. The glycerolysis reaction was done by heating the triglycerides in the presence of glycerol and a catalyst. In a typical reaction, glycerol was heated at 220–230°C for 2 h under an N_2 atmosphere to distill off any water present [33]. The amount

of soybean oil reacted with the glycerol was 4 g soybean oil to 1 g glycerol, a molar ratio of 1.75 mol glycerol to 1 mol triglyceride. The soybean oil was added in five portions to the glycerol, each portion 1 h apart. With the first portion, commercial soap was added in the amount of 1% of the total oil amount to act as an emulsifier and catalyst. The solution was heated at 230°C under N₂ while being stirred. After 5.5 h, the reaction was immediately cooled to room temperature with an ice bath, causing glycerol to separate from the

mixture. On removal of this layer, approximately 90% of the reaction solution, consisting of glycerides and glycerol, was recovered.

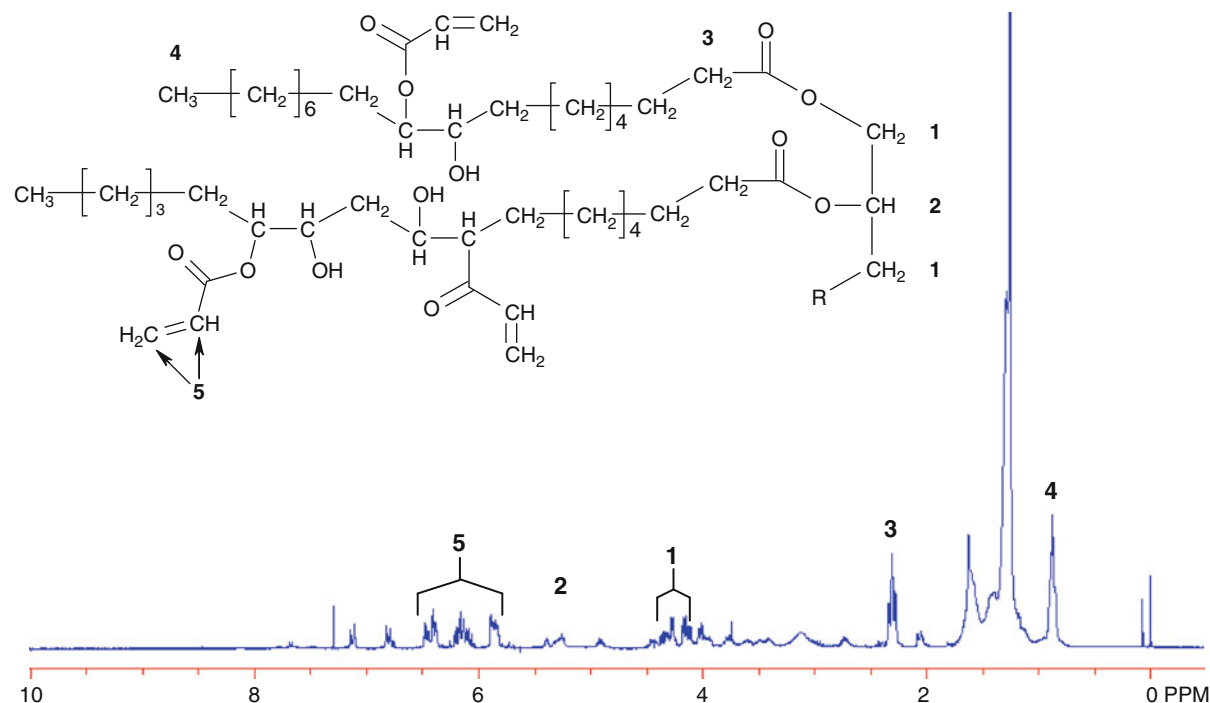
Maleinization of the mixture was accomplished by heating 60 g of glyceride/glycerol mixture to about 80°C while being stirred. Maleic anhydride was then added in the amount of 40 g. As the anhydride melted, 0.6 g triphenyl antimony was added as a catalyst along with 0.01 g hydroquinone. The reaction was complete after 5.5 h, according to FTIR and ¹H NMR, resulting in a mixture of maleinized glycerides and glycerol (SOMG/MA) [33].

New Polymers, Renewables as Raw Materials. Table 6
Reactant amounts used in malination of castor oil (CO:MA = 1:3)

Component	Weight(g)	Moles
Castor oil	100	0.108
Maleic anhydride	31.82	0.325
<i>N,N</i> dimethyl benzyl amine	1.32	–
Hydroquinone	0.132	–

Maleinized Soybean Oil Monoglyceride/Neopentyl Glycol Resin Synthesis

Modifying the procedure given in [53], SOMG/NPG/MA resin was synthesized as follows [51]. Forty-five grams of SOMG was placed into a 250-mL round-bottom flask equipped with a temperature controller and a magnetic stirrer and then heated to 125°C. Fifteen grams of NPG (0.144 mol) was then added to SOMG, and as the NPG melted, 58.3 g maleic



New Polymers, Renewables as Raw Materials. Figure 12

¹H-NMR spectrum for acrylated epoxidized soybean oil (Ebecryl 860, UCB Chemicals Co.)

anhydride was added. As the three compounds formed a homogenous solution, 0.06 g triphenyl antimony catalyst and 0.015 g hydroquinone were added. The solution was stirred for 6.5 h at 120°C. ^1H NMR analysis of the product showed the formation of both the SOMG and NPG maleate and later fumarate vinyl groups. The product was a light yellow viscous liquid at room temperature.

Maleinized Soybean Oil Monoglyceride/Bisphenol a Resin Synthesis

The preparation of maleates of BPA and ethylene and propylene oxide adducts of BPA was reported in [52]. For this work, SOMG and BPA were maleinized as a mixture [51]. Forty-five grams of SOMG was placed into a 250-mL round-bottom flask equipped with a temperature controller and a magnetic stirrer and heated to 125°C. Fifteen grams of BPA (0.0657 mol) was added to the SOMG, and as BPA dissolved, 42.88 g maleic anhydride (0.4375 mol) was added. As the three compounds formed a homogenous solution, 0.6 g triphenyl antimony and 0.01 g hydroquinone were also added. The solution was then stirred for 9 h at 125°C until maleic anhydride consumption was completed. The ^1H NMR analysis of the product showed the formation of both the SOMG and BPA maleate and later fumarate vinyl groups. The reaction product was an orange-colored viscous liquid (98 g) at room temperature.

Copolymerization of the Maleates with Styrene

The copolymerization of SOMG/MA, SOMG/NPG/MA, and SOMG/BPA/MA with styrene were all run under the same conditions for comparison of the mechanical properties of the resulting polymers. For this purpose, a certain weight ratio of the maleate mixture was mixed with 35% of its own weight of styrene in a closed vial. All of the maleate products were found to be soluble in styrene. *tert*-Butyl peroxybenzoate radical initiator, 2% by weight of the total mixture, was added. Nitrogen gas sparging and vacuum degassing were carried out for 5 min. The solution was then transferred to a rectangular rubber gasket mold sandwiched between two steel plates. The resin-filled mold was heated to 120°C at a rate of 5°C/min and was cured at this temperature for 3.5 h.

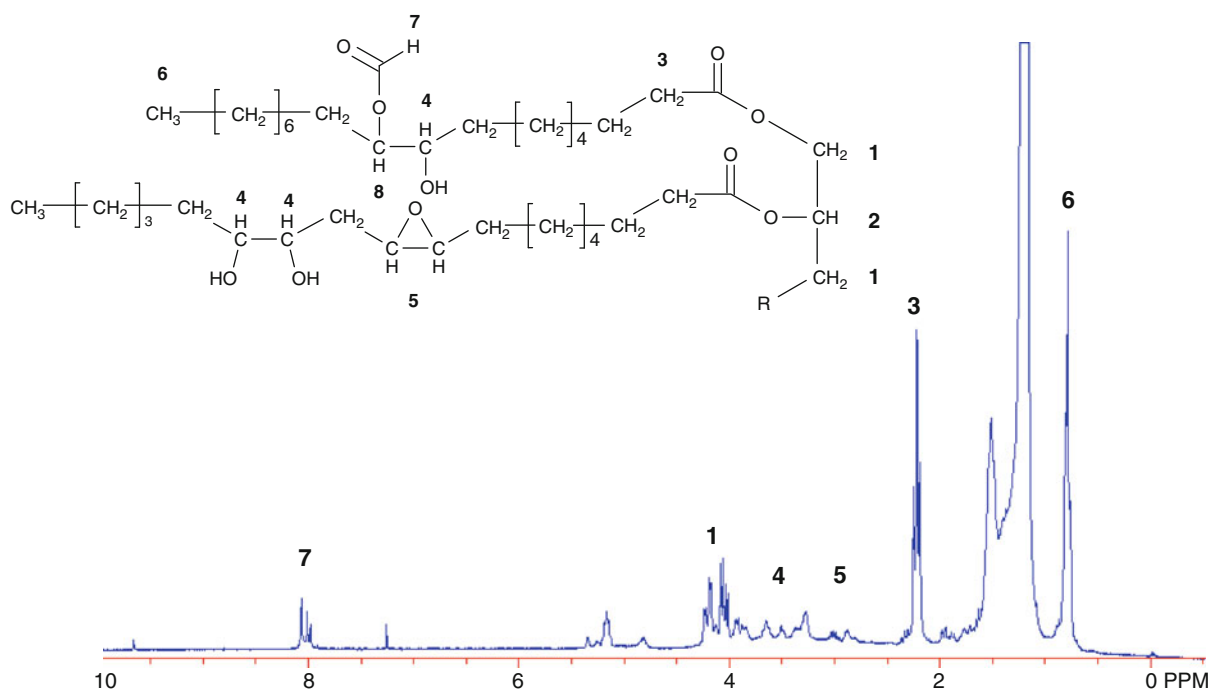
It was then postcured at 150°C for 1 h. Samples were clear, homogeneous, and free of voids or gas bubbles. The polymer samples were polished and prepared for DMA, which was conducted in a three-point bending geometry on a Rheometrics Solids Analyzer II. The temperature was ramped from 30°C to 200°C at a rate of 5°C/min, with a frequency of 1 Hz and strain of 0.01%.

Malinated Hydroxylated Oil Polymer Synthesis

The HO/MA shown in Fig. 4 uses the unsaturation of the triglyceride to incorporate polymerizable groups. This monomer was used in a series of experiments to understand how triglyceride structure can affect the synthesis and dynamic mechanical properties of the end polymer [59]. Olive oil, cottonseed oil, soybean oil, safflower oil, linseed oil, triolein, and a genetically engineered high oleic soybean oil were converted into HO/MA resins. The levels of unsaturation for these oils are shown in Table 1. The fatty acid chain lengths for all of these oils are between 17.5 and 18 carbons, making the unsaturation level essentially the only difference among oils.

Hydroxylation was done by stirring the oil (~100 g) vigorously in the presence of formic acid (150 mL) and 30% (aq) hydrogen peroxide (55 mL) at 25°C [30, 59]. The reaction time was 18 h to reach a maximum conversion of double bonds. Formic acid, peroxide, and water were then removed from the hydroxylated oil by dissolving the reaction mixture in diethyl ether and washing multiple times with water and then saturated (aq) sodium bicarbonate until a neutral pH was reached. The solution was then washed with saturated (aq) sodium chloride and dried over sodium sulfate. Finally, the ether was evaporated off under vacuum. The extent of hydroxylation can be characterized by ^1H NMR. An example ^1H NMR spectrum is presented in Fig. 13 with corresponding peak assignments [59]. The extent of hydroxylation has a linear dependence on the level of unsaturation. Generally, for every double bond present on the triglyceride, an average of 1.6 hydroxyls can be added [59].

The purified hydroxylated oil was reacted with maleic anhydride in a ratio of 1 mol triglyceride to 4 mol anhydride. The hydroxylated oil was heated to a temperature of about 80°C, and finely ground maleic



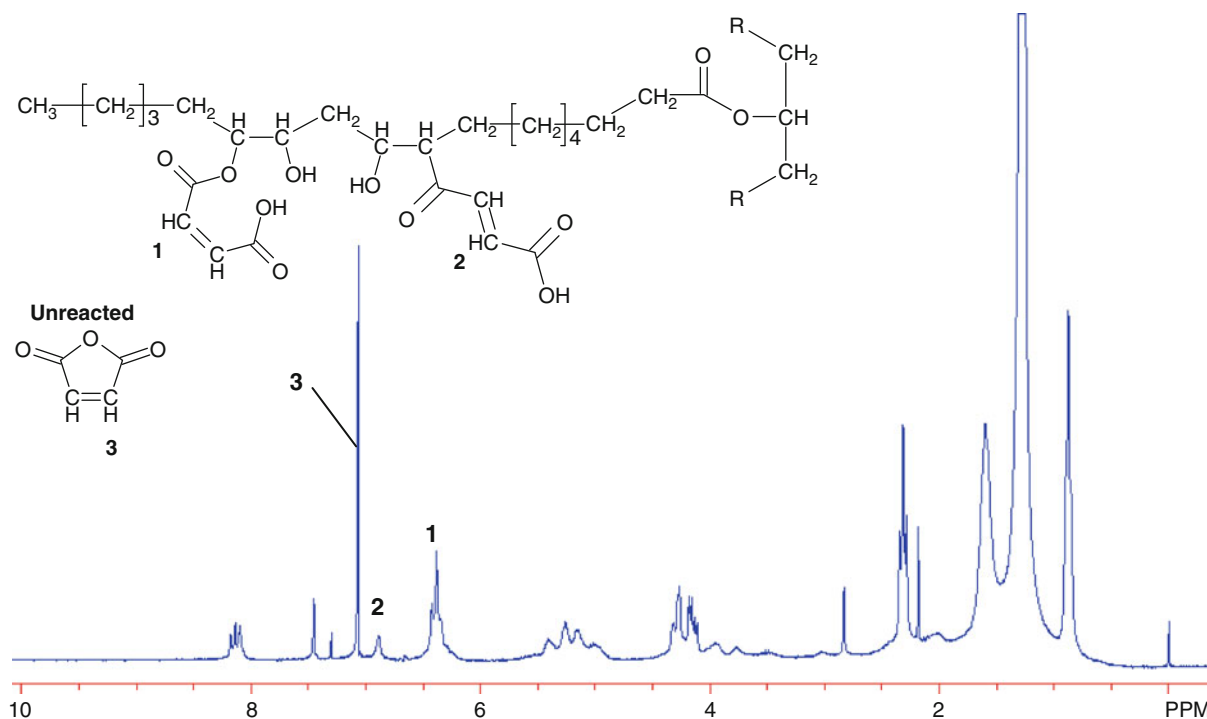
New Polymers, Renewables as Raw Materials. Figure 13

^1H -NMR of hydroxylated soybean oil. Treating the oil with formic acid and hydrogen peroxide results in conversion of the double bonds to hydroxy groups

anhydride was then added. Upon dissolving of the anhydride, *N,N*-dimethylbenzylamine was added to catalyze the reaction. The reaction was continued for 3 h, and the extent of maleinization was determined by ^1H NMR. An illustrative ^1H NMR is shown in Fig. 14 [59]. Under these reaction conditions, the extent of functionalization plateaus in the range of 2.1–2.8 maleates/triglyceride for all oils. Approximately 20–25% of the maleates attached to the triglycerides isomerize to form fumarate groups (*trans* confirmation). Unreacted maleic anhydride remained in the resin and was polymerized during the cure reaction. The HO/MA resins were then dissolved in styrene in a molar ratio of 7:1 styrene to HO/MA. Resins were cured using 2,5-dimethyl-2,5-di-(2-ethylhexanoylperoxy)hexane at 65°C for 1.5 h and postcured at 120°C for 1 h. DMA was conducted in a three-point bending geometry on a Rheometrics Solids Analyzer II. Temperature was ramped from 30°C to 175°C at a rate of 5°C/min, with a frequency of 1 Hz and strain of 0.01%.

SOPERMA and COPERMA Polymer Synthesis

The general-purpose unsaturated polyester (UP) resin is a linear polymer with the number average molecular weights in the range of 1,200–3,000 g/mol. Depending on the chemical composition and molecular weight, they can be viscous liquids or solids. The plant oil-based resins prepared are not linear polymers but similar mixtures of monomers or oligomers with different molecular weights. The number and weight average molecular weights of different species in each of these plant oil-based resins were in the range of ~300–2,000 g/mol as presented in section “[Synthetic Pathways for Triglyceride-Based Monomers](#).” With the exception of the COMA and SOGLYME resins, which were liquid at room temperature, all the malinated glyceride-based resins prepared were paste-like solids at room temperature. The melting points of these resins were in the range of 60–70°C, similar to the general-purpose UP resins whose melting points are in the range of 60–77°C.



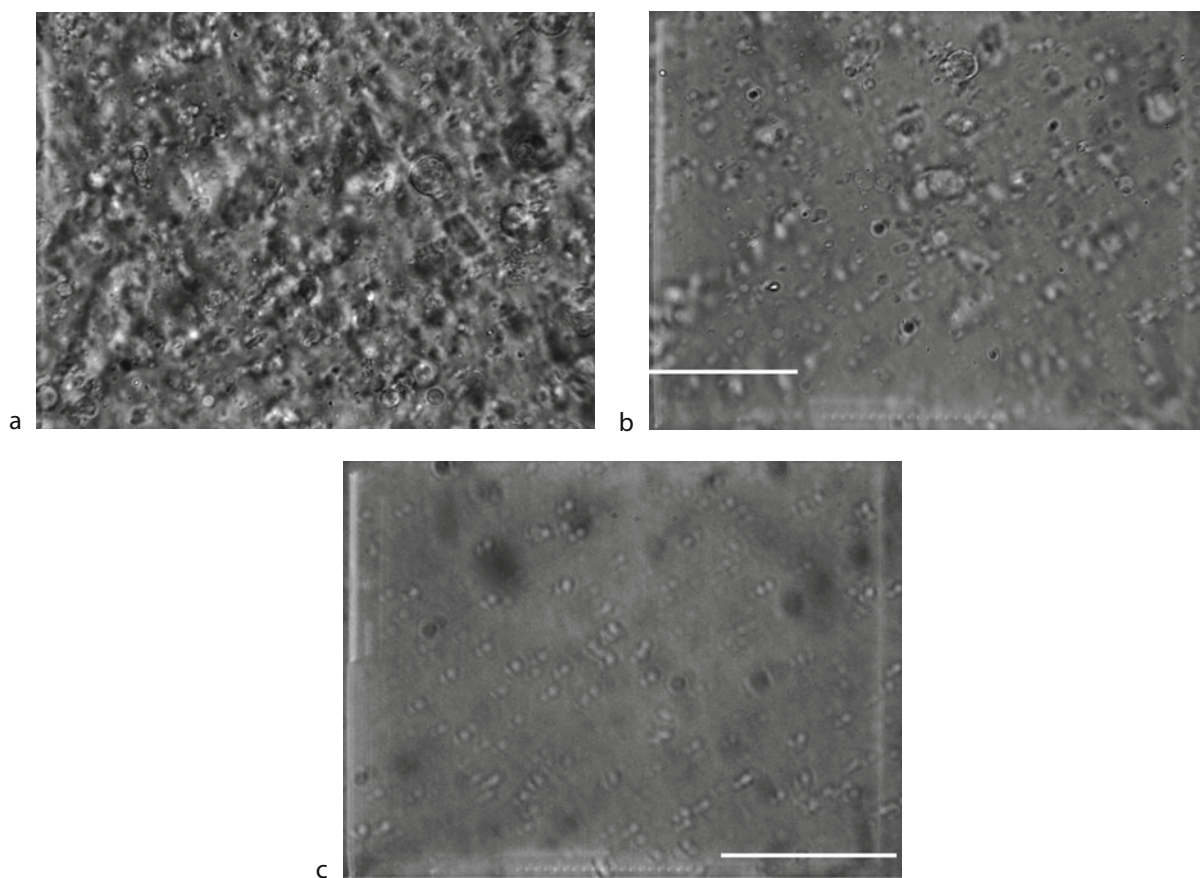
New Polymers, Renewables as Raw Materials. Figure 14

¹H-NMR of maleinized hydroxylated soybean oil. Peaks 1 and 2 represent the maleate half esters and fumarate half esters, respectively. Peak 3 represents unreacted maleic anhydride

To prepare the styrenated plant oil-based resins, which were solid at room temperature, the resin was first heated in an oil bath above its melting point ($\sim 70^{\circ}\text{C}$), then the necessary amount of styrene was added and the mixture was agitated in the oil bath at a temperature of 80°C until styrene and the resin became totally mixed. The resin was then processed at room temperature for polymerization. For the resins that were liquid at room temperature, styrene was added, mixed, and processed at room temperature. Apart from the SOGLYME resin (methacrylated soybean oil monoglycerides), all the styrenated resins showed phase separation in the microscale. Figure 15 shows the optical microscopic pictures of the SOPERMA-styrene mixtures at decreasing SOPERMA concentrations at $500\times$ magnification. The SOPERMA forms nonuniform droplets in the continuous styrene matrix in the $1\text{--}5\text{-}\mu\text{m}$ size range. As can be seen in Fig. 15, these droplets become less dense and more uniform in size as the concentration of SOPERMA decreases from 80 to 20 wt%. Similarly, COPERMA,

COGLYMA, and COBPAPRMA resins containing 33 wt% styrene were found to be incompatible with styrene and exhibited phase separation.

The general-purpose UP resins based on propylene glycol, phthalic, and maleic anhydride are miscible with styrene. The molecular weights of the general-purpose UP resins have a number average molecular weight of 900 g/mol and a weight average molecular weight of 2,400 g with a polydispersity of 2.7. The acid number of these polyesters is around 50 mg KOH/g. The incompatibility of the malinated plant oil-based resins in styrene can be attributed to both the abundance of the acid groups of the maleate half-esters and the presence of high-molecular-weight species in these resins. The acid number of all the malinated glyceride-based resins was found to be above 200 mg KOH/g. This value is much higher than that of the general-purpose UP resins. The molecular species present in the malinated glyceride-based resins are highly polar compared to styrene, and the strong interactions of these molecules via hydrogen bonding between the acid groups result in



New Polymers, Renewables as Raw Materials. Figure 15

Optical microscopic pictures of the SOPERMA-styrene mixtures (500 \times) (a) 80 wt% (b) 60 wt% (c) 20 wt% SOPERMA

the insolubility of the malinated glycerides in styrene. Thus, the SOGLYME resin based on methacrylated glycerides, both with the acrylic acid by-product and without acrylic acid, showed no phase separation in styrene. The methacrylate half-esters as shown earlier in Fig. 8 do not carry acid functionality, and oligomer formation during the methacrylation reaction cannot occur. Additionally, these molecules are less polar than maleate half-esters and they cannot interact via hydrogen bonding, which favors their compatibility with styrene. The SOPERMA resin was also found to be insoluble in more polar monomers such as acrylic acid and acrylonitrile. The immiscibility of the SOPERMA resin in these solvents shows that the molecular weight and strong interactions between the malinated glyceride molecules play an important

role in the immiscibility of these molecules in polar solvents.

A liquid molding resin should have properties within a certain operating range to be successfully used in molding processes. Possibly the most stringent requirement is the resin's viscosity, which must range between 200 and 1,000 cP. At viscosities lower than 200 cP, air pockets will remain in the mold after injection. At viscosities greater than 1,000 cP, voids may occur in the part, the time required for injection increases, and problems with fiber wetting can arise during composite preparation.

The surface free energy of a liquid, also referred to as *surface tension*, determines most of the surface and interfacial properties such as wetting, adhesion, and adsorption. Surface tension results from an imbalance

of molecular forces in a liquid. At the surface of the liquid, the liquid molecules are attracted to each other and exert a net force, pulling themselves together. High surface tension values mean that the molecules tend to interact strongly; thus, polar materials show high surface energy values. The surface energy of a liquid molding resin may be especially important for reinforcement of the resins by fibers. The wetting of a fiber with a liquid resin can be judged by the difference between the surface energies of the fiber and the resin. For the most desirable condition, proper wetting and spreading in resin transfer molding processes, the surface energy of the fiber should be high, whereas the surface energy of the resin should be low. Table 7 shows the specific gravity, viscosity (η), and surface energy values for the plant oil-based resins and general-purpose UP resins containing 33 wt% styrene.

As can be seen from Table 7, apart from the COMA and the SOGLYME resins, these resins show viscosities that are in a range that is suitable for liquid molding processes. Although the COMA and SOGLYME resins show low viscosities, the viscosities of these resins can easily be increased by decreasing the amount of the styrene diluent. It was found by Can [66] that as the weight fraction of styrene was increased, the viscosity decreased in an exponential manner. This result is very desirable since it shows that a small amount of

comonomer can be used to make these resin systems much easier to process.

The surface energy values of these resins show values closer to those of UP resins, and are significantly lower than those of vinyl ester resins (32–34 mN/m), which may have difficulty in wetting fiber substrates. Among the malinated resins, the surface free energy value is highest for the COPERMA resin and lowest for the COMA resin in proportion to the maleate content of the resins. A higher maleate content should result in a higher polarity and a higher surface free energy. The SOGLYME resin shows the lowest surface free energy among all the resins. The methacrylated glycerides do not carry an acid functional group as malinated glycerides and, therefore, are less polar than the corresponding maleates.

The curing of UP resin is accomplished via free-radical cross-linking polymerization between the UP molecules and styrene. The UP molecules are the cross-linkers, while the styrene acts as an agent to cross-link the adjacent polyester molecules. Similarly, in the plant oil-based resin systems, the functionalized glycerides act as the cross-linker units, and styrene is the agent that links the adjacent glyceride molecules. Styrene is the most commonly used vinyl monomer in unsaturated polyester resins due to its low viscosity, low cost, and reactivity with the unsaturated sites of polyesters. The unsaturation present on the UP backbone is very sluggish in homopolymerization. The reactivity ratio of styrene and maleic/fumaric acid esters is about 0, indicating that this system has a tendency to form alternating copolymers. Because the malinated plant oil-based resins were found to be insoluble in styrene at room temperature, it was especially important for us to determine the conversion of polymerization for the maleate and styrene monomers.

For the determination of the conversion and kinetics of the polymerization, the styrenated SOPERMA, COMA, and COPERMA resins were prepared as described earlier. The styrene concentration was 33 wt% for each resin. *tert*-Butyl peroxybenzoate was used as the initiator. The initiator concentration was 2 wt% for the SOPERMA and COMA resins and 1.5 wt% for the COPERMA resin. The curing of all resins was carried out at 120°C for 3 h for comparison of conversion and rate of polymerization in different resins. The SOPERMA and COPERMA resins were also postcured

New Polymers, Renewables as Raw Materials. Table 7
The specific gravity, viscosity, η and surface energy values for the plant oil based resins (33 wt% styrene) and general purpose UP resins

Resin	Specific gravity (g/ml)	Viscosity, η (cP)	Surface energy (mN/m)
SOPERMA	0.94	343	27.38
COPERMA	1.06	363	28.84
COGLYCMA	1.04	213	26.36
COBPAPRMA	0.98	183	27.2
COMA	0.75	92	26.02
SOGLYME	0.87	51	22.7
GP-UP	1.14	200–2,000	24–30

at 150°C and 160°C, respectively, for 1 h after 2 h at 120°C. Although, all the resins showed similar conversion versus time profiles during the isothermal cure, the final conversion values were different for each resin.

Table 8 lists the total conversion (α) and the conversion of maleates (α_{ma}) and styrene (α_{st}) for the COMA, SOPERMA, and COPERMA resins (33 wt% styrene) at the end of 3 h at 120°C. As can be seen, the final maleate conversion decreases as the maleate content of the resin increases from COMA to COPERMA. The styrene conversion and thus the total conversion is also considerably lower for the SOPERMA and COPERMA resins than for the COMA resin. It is expected that the molecular mobility of the resin decreases as the cross-link density increases, resulting in lower total conversion. As a result, styrene monomer as well as some resin can be trapped in the network and cannot participate in polymerization. However, although the COPERMA resin has a higher maleate content and, therefore, a higher cross-link density than the SOPERMA resin, the styrene and thus the final conversions for the SOPERMA and COPERMA resins do not show a significant difference. The maleate conversion was higher than styrene conversion for all of the resins.

The total ultimate conversion of unsaturated polyesters ranges from 0.75 to 0.9 and increases with increasing temperatures. Similarly, the final conversions for the isothermal cure of the plant oil-based resins were lower than the complete conversion. During an isothermal cure, when the increasing glass transition temperature of the resin reaches the reaction temperature and the material evolves from the rubbery state to the glassy state, the rate of propagation

becomes diffusion controlled. This process, referred to as *vitrification*, may virtually terminate the polymerization, limiting the conversion that can be reached isothermally. This was true in particular for the SOPERMA and COPERMA resins, which possess glass transition temperatures (T_g) when fully cured [$T_g(\text{SOPERMA}) = 139^\circ\text{C}$ and $T_g(\text{COPERMA}) = 146^\circ\text{C}$] that are higher than the cure temperature. Thus a postcure was necessary for these systems to increase the conversion.

Properties of Plant Oil Resins

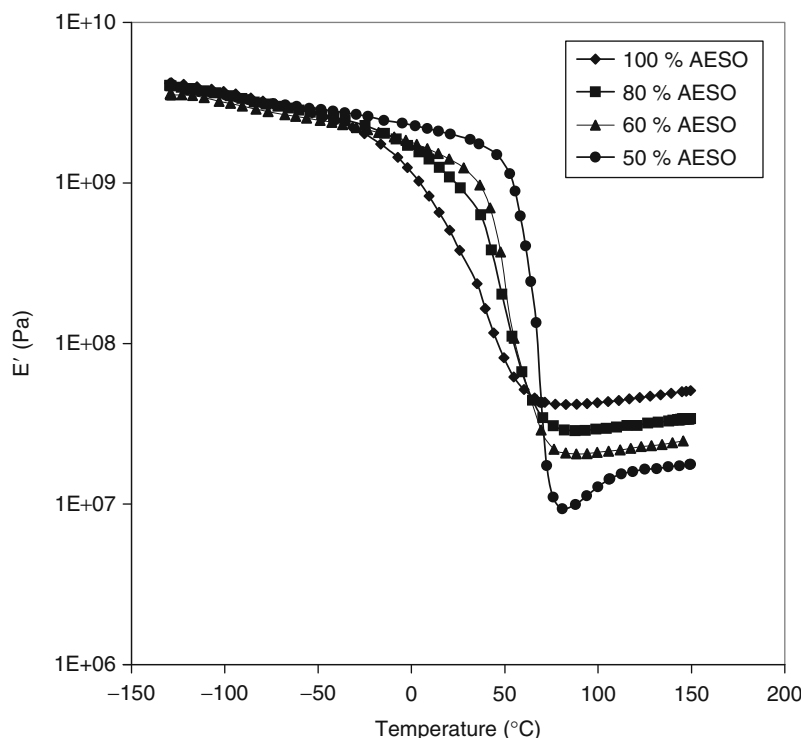
Viscoelastic and Mechanical Properties of AESO-Styrene Polymers

The storage moduli, E' , of the AESO-styrene neat polymers at various temperatures and compositions are shown in Fig. 16. At room temperature, the polymers display moduli proportional to the amount of styrene present, which is expected from the tensile properties presented earlier. Additionally, at room temperature all of the polymers are in the transition phase from the glassy region to the rubbery plateau. Even at temperatures as low as -130°C , it does not appear that these polymers have reached a characteristic glassy plateau. At extremely low temperatures, all compositions exhibit essentially equal moduli of about 4 GPa. At higher temperatures, the compositions show moduli inversely proportional to the amount of styrene present. According to rubber elasticity theory [60], the lower styrene content polymers have a higher cross-link density, as observed in Fig. 16.

The T_g is often designated by either the temperature at which the dynamic loss modulus E'' value is at a peak or the temperature at which the loss tangent $\tan \delta$ exhibits a peak [61]. As shown in Fig. 17, all of the AESO-styrene copolymers exhibit two peaks in E'' . A minor relaxation occurs in the range of -85°C to -95°C , showing little dependence on composition. The much larger relaxation, corresponding to the T_g , occurs in the range of -10°C to 60°C and also becomes sharper in nature with the addition of styrene. These peaks are shown in the $\tan \delta$ graph in Fig. 18. The temperature at which these peaks occur exhibits a linear dependency on composition, increasing with the amount of styrene present in the system, as illustrated by Fig. 19.

New Polymers, Renewables as Raw Materials. Table 8
The total conversion(α), the conversion of maleates, (α_{ma}), and styrene (α_{st}), for the SOPERMA, COMA and COPERMA resins (33 wt% styrene) at the end of 3 hours at 120°C

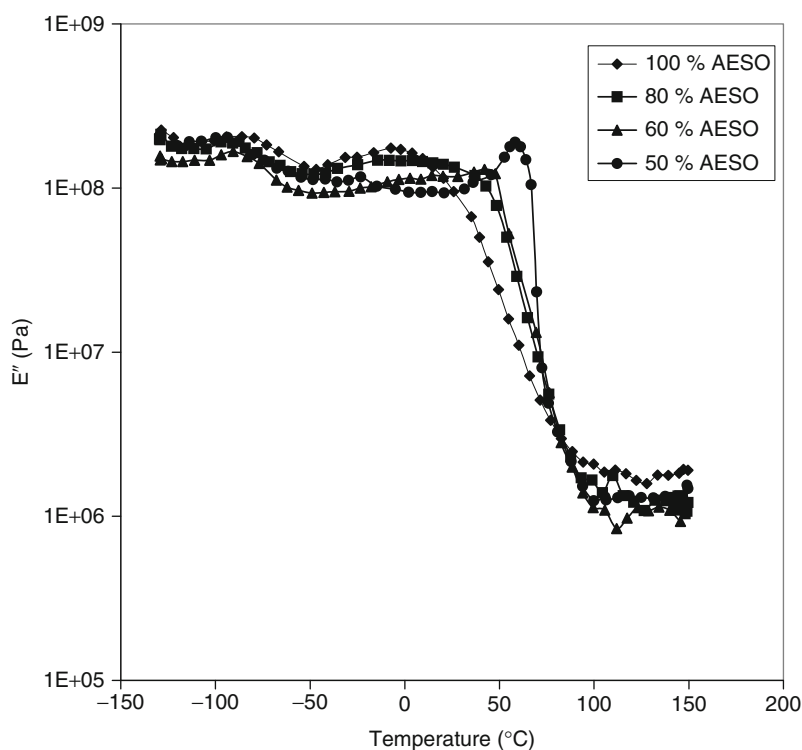
Resin	α_{ma}	α_{st}	α
COMA	0.998	0.921	0.967
SOPERMA	0.979	0.828	0.886
COPERMA	0.952	0.835	0.885



New Polymers, Renewables as Raw Materials. Figure 16
Storage modulus (E') of AESO-styrene copolymer as a function of temperature

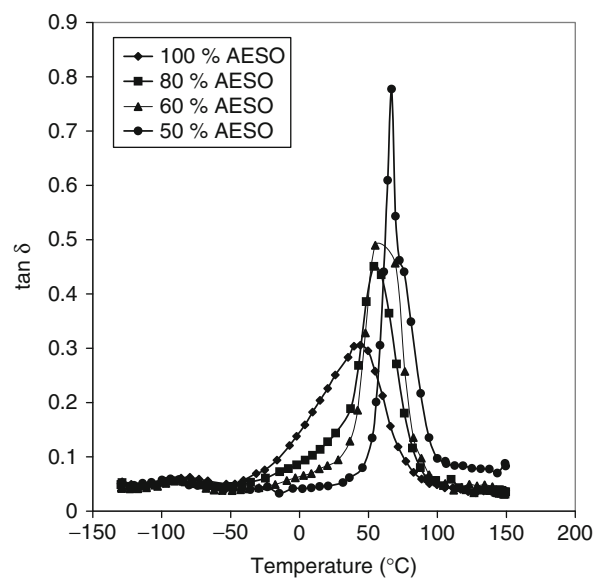
The dynamic mechanical behavior just discussed is a combination of three factors: cross-link density, copolymer effects, and plasticization. As the amount of AESO increases, so does the number of multifunctional monomers. Therefore, the overall cross-link density will be greater with increasing amounts of AESO, as supported by the high-temperature moduli shown in Fig. 16. Increasing the cross-link density slows the transition in E' from glassy to rubbery behavior. Additionally, the $\tan \delta$ peak broadens and decreases in height [61]. The copolymer effect occurs frequently when there are differences in the reactivity or structure of the different monomers. If one monomer is more reactive, it is depleted faster, causing polymer formed later in the reaction to be composed mostly of the slower reacting monomer. This causes heterogeneity in the composition of the total polymer. If these monomers differ from each other in their physical properties, such as very different T_g 's, a general broadening of the glass-rubbery transition is frequently observed, due to this gradient [61].

The other factor in the dynamic mechanical behavior, plasticization, is due to the molecular nature of the triglyceride. The starting soybean oil contains fatty acids that are completely saturated and cannot be functionalized with acrylates. Therefore, these fatty acids act in the same manner as a plasticizer, introducing free volume and enabling the network to deform more easily. The addition of even small amounts of plasticizer to polymers was known to drastically broaden the transition from glassy to rubbery behavior and reduce the overall modulus [61]. This plasticizer effect presents an issue that may be inherent to all natural triglyceride-based polymers that use the double bonds to add functional groups. However, with advances in genetic engineering, it may be possible to reduce this trend by reducing the amount of saturated fatty acids present, thus sharpening the glass-rubbery transition. This issue is addressed later in the properties of HO/MA polymers produced from genetically engineered high oleic content oil and synthetic triolein oil. The existence of some saturated fatty acids, though,



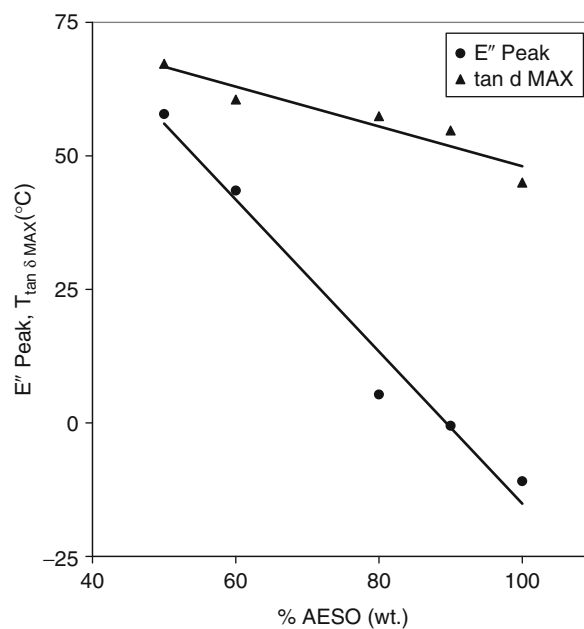
New Polymers, Renewables as Raw Materials. Figure 17

Loss modulus (E'') of AESO-styrene copolymer as a function of temperature



New Polymers, Renewables as Raw Materials. Figure 18

Damping peak ($\tan \delta$) of AESO-styrene copolymer as a function of temperature



New Polymers, Renewables as Raw Materials. Figure 19

E'' and $\tan \delta$ peak temperatures of various compositions of AESO-styrene polymer

can contribute to improved toughness and ballistic impact resistance [62].

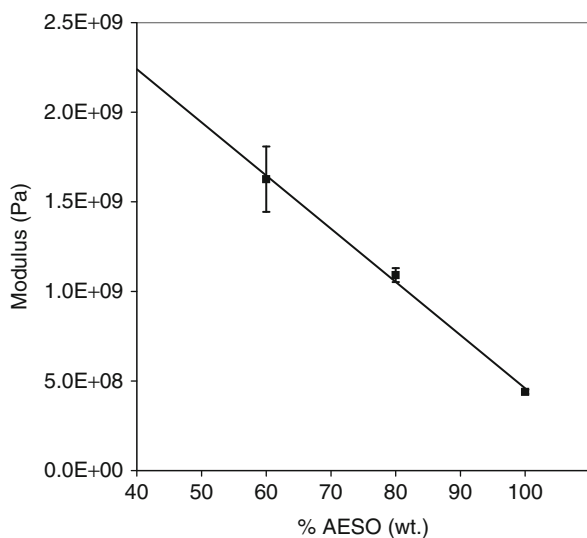
Tensile Properties of AESO-Styrene Polymers

The tensile moduli of three AESO-styrene copolymers at room temperature are shown in Fig. 20. The pure AESO polymer has a modulus of about 440 MPa. At a styrene content of 40 wt%, the modulus increases fourfold to 1.6 GPa. In this region, the dependency on composition appears to be fairly linear. The ultimate tensile strengths of these materials, as shown in Fig. 21, also show linear behavior. The pure AESO samples exhibited strengths of approximately 6 MPa, whereas the polymers with 40 wt% styrene show much higher strengths of approximately 21 MPa. Therefore, it is apparent that the addition of styrene drastically improves the properties of the end resin.

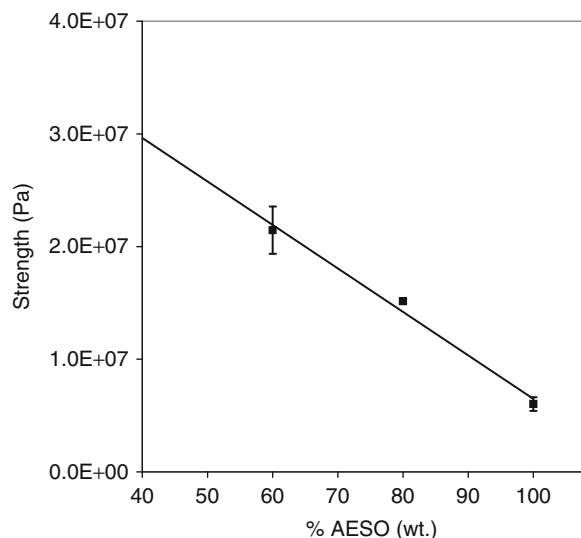
Dynamic Mechanical Behavior of Modified AESO Resins

The dynamic mechanical properties of the AESO polymers modified by cyclohexanedicarboxylic anhydride (CDCA) and maleic acid were better than the unmodified polymers. As shown in Fig. 22, the storage modulus at room temperature increases with both of

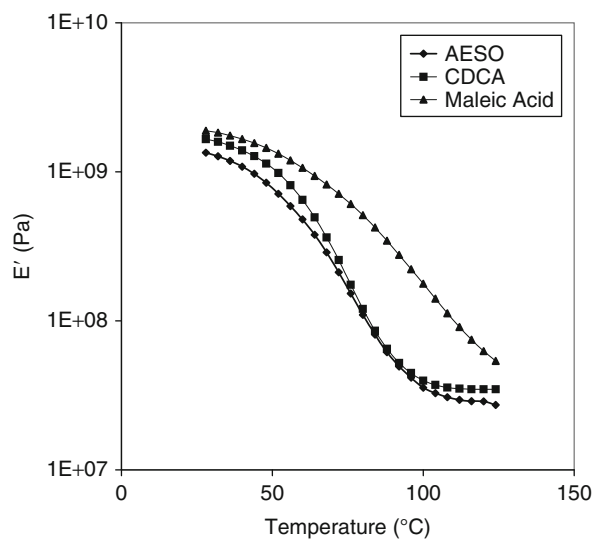
these modifications. The storage modulus of the unmodified AESO resin at room temperature is 1.3 GPa, whereas the cyclohexanedicarboxylic anhydride modification increases the modulus to 1.6 GPa. The maleic acid modification provides the most



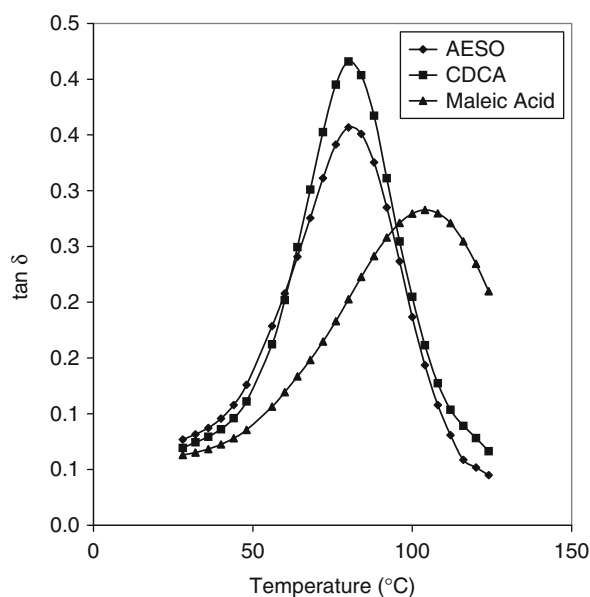
New Polymers, Renewables as Raw Materials. Figure 20
Tensile modulus of AESO-styrene copolymers



New Polymers, Renewables as Raw Materials. Figure 21
Ultimate tensile strength of AESO-styrene copolymers



New Polymers, Renewables as Raw Materials. Figure 22
Storage modulus (E') of modified AESO resins as a function of temperature



New Polymers, Renewables as Raw Materials. Figure 23 Damping peak ($\tan \delta$) of modified AESO resins as a function of temperature. Peaks in $\tan \delta$ were found at 81°C (CDCA modified) and 105°C (maleic acid modified) compared to 79°C for the synthesized AESO

improvement, raising the storage modulus to 1.9 GPa. The T_g , as indicated by the peak in $\tan \delta$, does not show any large increase from the anhydride modification, as shown in Fig. 23. However, the maleic acid modification shifts the $\tan \delta$ peak by almost 40 °C, showing a peak at 105 °C. The increased broadness of the peak can be attributed to increased cross-link density.

SOMG/MA Polymer Properties

As seen in Fig. 24, the $\tan \delta$ peak for the SOMG/MA polymer occurs at around 133°C, and the polymer has an E' value of approximately 0.92 GPa at room temperature. It is apparent that the glass transition is rather broad due to the broad molecular weight distribution of the SOMG maleates. The distribution of soybean oil monoglyceride monomaleates, monoglyceride bismaleates, diglyceride monomaleates, and glycerol tris maleates was confirmed by mass spectral analysis, which was reported in a previous publication [63]. The tensile tests performed on the copolymers of SOMG

maleates with styrene showed a tensile strength of 29.36 MPa and a tensile modulus of 0.84 GPa as calculated from the force displacement graph.

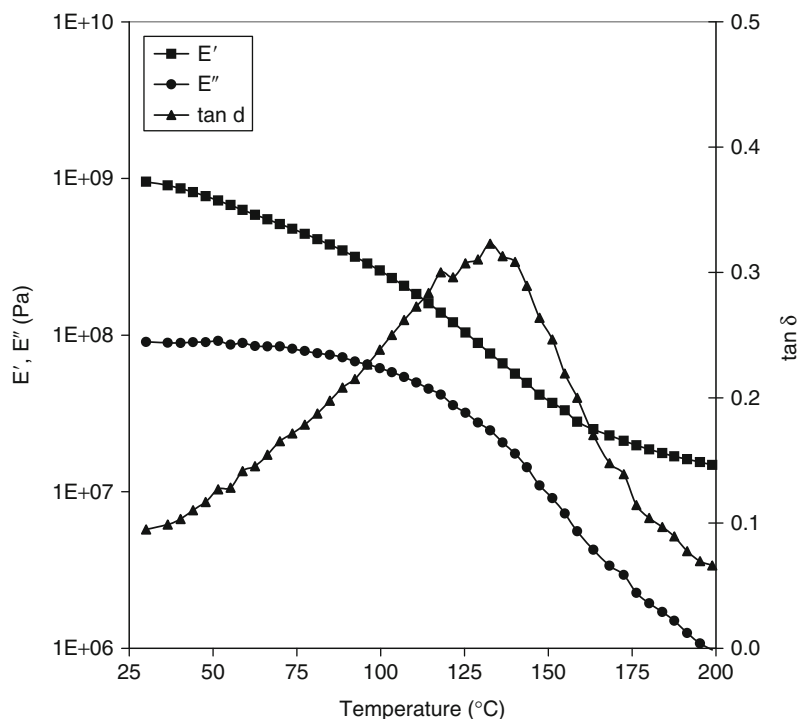
SOMG/NPG Maleates (SOMG/NPG/MA)

The DMA of SOMG/NPG/MA polymer showed a $\tan \delta$ peak at approximately 145°C and an E' value of 2 GPa at room temperature. The 12°C increase in the T_g and the considerable increase in the modulus of the copolymers of SOMG/NPG maleates with styrene compared to that of the SOMG maleates can be attributed to the replacement of the flexible fatty acid chains by the rigid methyl groups of NPG. The overall dynamic mechanical behavior of the SOMG/NPG/MA polymer was very similar to that of the SOMG/MA shown in Fig. 21. Despite the higher T_g and modulus, there remained a broad glass transition. The tensile strength of the SOMG/NPG/MA polymer was found to be 15.65 MPa, whereas the tensile modulus was found to be 1.49 GPa.

Maleinized pure NPG polymerized with styrene (NPG/MA) was prepared to compare its properties with the SOMG/NPG/MA polymer [51]. DMA analysis of the NPG/MA showed a $\tan \delta$ peak at approximately 103°C and an E' value of about 2.27 GPa at 35°C. The high T_g value observed for the SOMG/NPG/MA system (~145°C) is attributed to a synergistic effect of both the NPG and SOMG together since the T_g value observed for the NPG/MA system (~103°C) is much lower. This is probably due to the incorporation of the fatty acid unsaturation into the polymer in the SOMG/NPG/MA system. On the other hand, the comparatively higher E' value observed for the NPG maleates explains the increase in the E' observed for the SOMG/NPG/MA system compared to that of the SOMG/MA system. The decrease in tensile strength of the SOMG/NPG/MA system compared to that of SOMG/MA may be attributed to a broader molecular weight distribution of this system compared to that of the SOMG maleates.

SOMG/BPA Maleates (SOMG/BPA/MA)

The DMA of this polymer showed a $\tan \delta$ peak at around 131°C and an E' value of 1.34 GPa at 35°C. The introduction of the rigid benzene ring on the polymer backbone considerably increased the modulus of the final polymer compared to that of the SOMG



New Polymers, Renewables as Raw Materials. Figure 24
Dynamic mechanical behavior for SOMG/MA polymer

maleates. The T_g of this polymer, however, was not very different from that of the SOMG maleates (133°C). This was attributed to a lower yield in the maleinization of the BPA, as determined from ^1H NMR data [51]. Like the SOMG/NPG/MA polymer, the SOMG/BPA/MA displayed the characteristic gradual glass transition shown in Fig. 21.

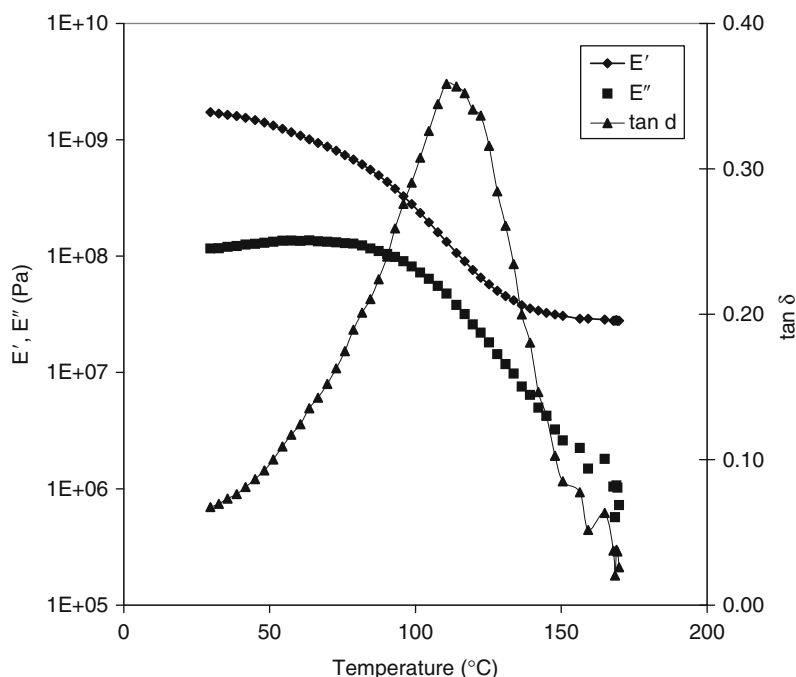
HO/MA Dynamic Mechanical Polymer Properties

The dynamic mechanical properties of the HO/MA polymers were found to be better than those of the AESO polymers. Little variation was seen between the polymers made from the different oils. At room temperature, the E' for all of the oils existed between 1.45 and 1.55 GPa, showing no dependence on saturation level. The dynamic mechanical behavior was similar between the different oils, with the typical behavior being shown in Fig. 25. The temperatures at which a maximum in $\tan \delta$ was exhibited ranged from 107°C to 116°C, which are all substantially higher than the AESO base resin. These properties are fairly

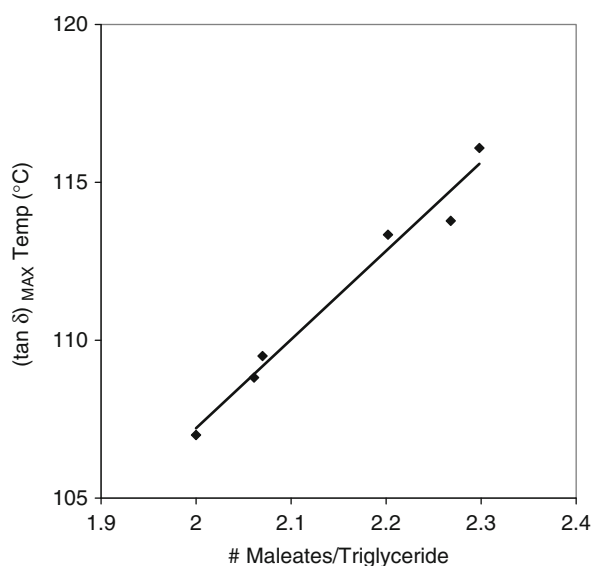
close to those shown by conventional petroleum-based polymers. However, the distinctive triglyceride behavior still exists, in that the glass transitions are extremely broad and that, even at room temperature, the materials are not completely in a glassy state. Again this is probably due to the saturated fatty acids of the triglycerides that act as a plasticizer.

Although the extent of maleinization was approximately the same from oil to oil, it is possible to see how the slight differences affect the T_g . In Fig. 26, the $\tan \delta$ peak temperature was plotted as a function of maleate functionality. Within this range, the behavior is linear, suggesting that if higher levels of functionalization are able to be reached, the properties should improve accordingly [64]. However, it is expected that past a certain extent of maleate functionality, the $\tan \delta$ peak temperature dependence will plateau. Work is currently being pursued to test the limits of this behavior.

It was previously stated that the broadness in the glass transition may be inherent to all triglyceride-based polymers. However, work with genetically engineered oil and synthetic oil has shown that it is



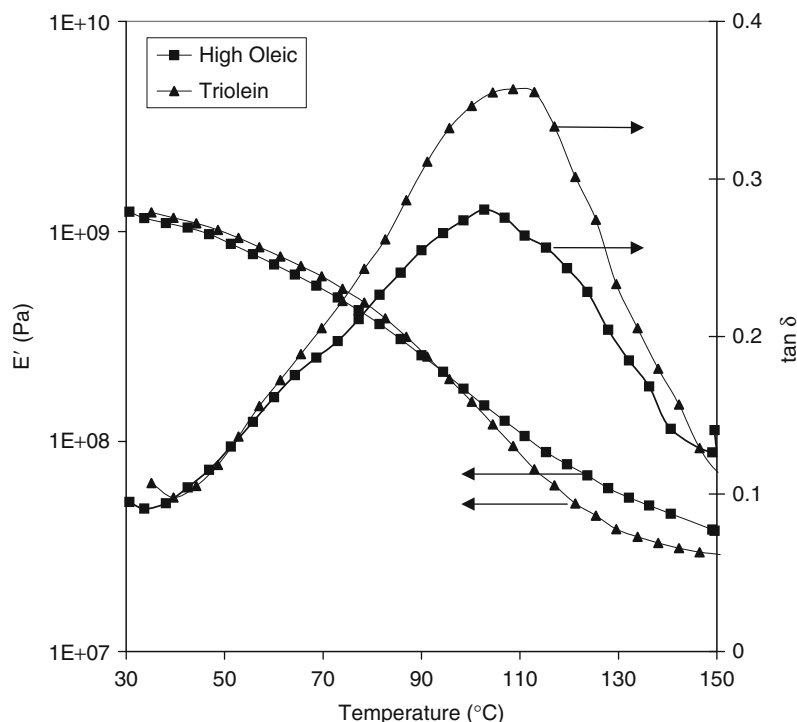
New Polymers, Renewables as Raw Materials. Figure 25
Representative dynamic mechanical behavior for HO/MA polymers



New Polymers, Renewables as Raw Materials. Figure 26
Peak in tan δ as a function of maleate functionality

possible to reduce this characteristic. The genetically engineered high oleic soybean oil has an average functionality of three double bonds/triglyceride and has the fatty acid distribution shown in Table 1. The maleinized form of this oil had a functionality of two maleates/triglyceride. The properties of polymers from this material were compared to polymers from triolein oil, which is monodisperse, consisting only of oleic fatty acid esters (18 carbons long and one double bond). The maleinized triolein oil had a functionality of 2.1 maleates/triglyceride. Thus, the only difference between the two oils is the fatty acid distribution of the high oleic oil versus the monodisperse triolein oil.

The dynamic mechanical properties of polymers made from these oils are shown in Fig. 27. The T_g of these two polymers does not seem to differ much, judging from either their tan δ peak or the inflection in the E' . However, the broadness of the transitions does differ. It is apparent that the triolein polymer has



New Polymers, Renewables as Raw Materials. Figure 27

Dynamic mechanical properties of polymers made from maleinized hydroxylated high oleic oil and triolein oil. The monodisperse triolein displays a sharper transition from the glassy region to the rubbery region

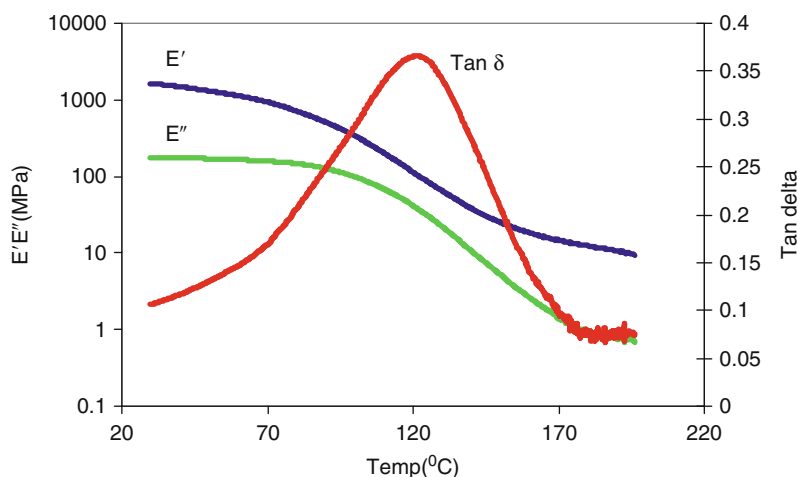
a sharper E' transition from the glassy region to the rubbery region. This is evident also in the $\tan \delta$ peak, which has a higher peak height. The transition is not yet as sharp as petroleum-based polymers. This is probably caused by the triolein monomer having a functionality of only two maleates/triglyceride. Consequently, there is still a plasticizer effect present, but this effect may be reduced by controlling the reaction conditions to reach higher conversions.

SOPERMA Polymer Properties

The typical DMA behavior of the SOPERMA-styrene polymer (40 wt% styrene) is shown in Fig. 28, where it can be seen that, at room temperature, these polymers are already in transition from the glassy region to the rubbery plateau. Most thermoset polymers show a distinct glassy region in which the modulus is independent of temperature. This is not observed for the SOPERMA-styrene polymers. The SOPERMA-styrene polymers show a very broad transition from the glassy

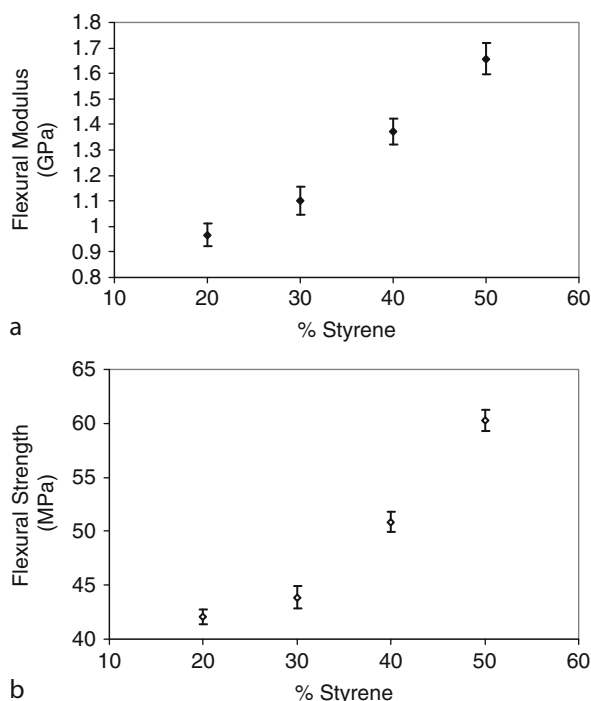
to the rubbery state. Because of this broad transition, these polymers do not show a clear peak in the loss modulus E'' . Thus, the $\tan \delta$ curve is also very broad. The broad transition observed for the SOPERMA-styrene polymers is a result of two major effects. One major effect is the phase separation, which results in higher T_g SOPERMA-rich and lower T_g styrene-rich regions in the polymer matrix. Another effect that may result in a broad glass transition is the plasticizing effect of the fatty acids present in the SOPERMA monomer which are not functionalized. The transition from the glassy to the rubbery state broadens significantly with the addition of small amounts of plasticizers to polymers.

Figures 29a and b show the flexural modulus and flexural strength of the SOPERMA-styrene polymers as a function of styrene concentration. As can be seen both the flexural modulus and flexural strength of the polymers increase with increasing concentrations of styrene despite the decrease in cross-link density, v . Thus, the rigid aromatic structure of the styrene monomer



New Polymers, Renewables as Raw Materials. Figure 28

Typical DMA behavior of the SOPERMA-styrene polymer (40 wt% styrene)



New Polymers, Renewables as Raw Materials. Figure 29
The change of (a) flexural modulus (b) flexural strength of SOPERMA-styrene polymers at increasing styrene weight percentages

as compared to the SOPERMA monomer with flexible fatty acid chains dominates the effect of cross-link density. The linear dependence of flexural modulus on styrene concentration above 30 wt% styrene follows Eq. 1:

$$E_f = 0.0278(W_{\text{styrene}}) + 0.2643, \quad (1)$$

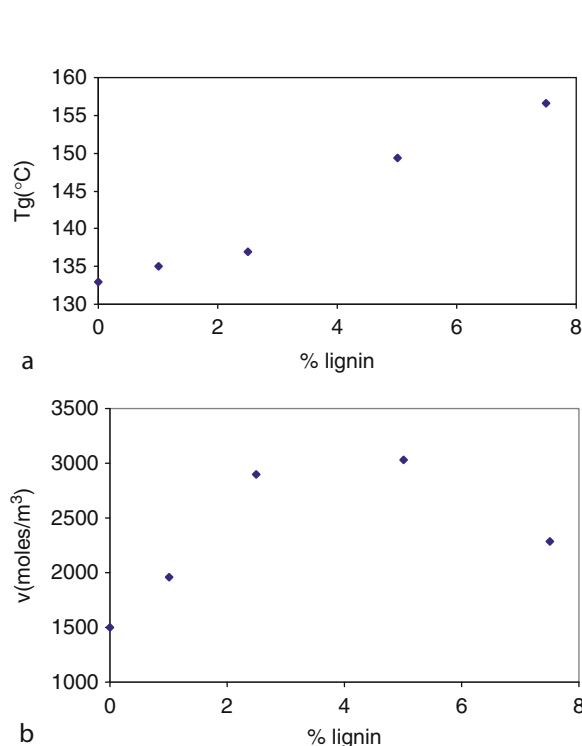
where W_{styrene} presents the weight percentage of styrene. This correlation predicts the flexural modulus of 100% polystyrene as 3.04 GPa. The flexural modulus values of standard polystyrenes of different grades are in the range of 2.9–3.8 GPa. In the same manner, the dependence of flexural strength on styrene concentration follows Eq. 2:

$$S_f = 0.6159(W_{\text{styrene}}) + 27.696 \quad (2)$$

which predicts the flexural strength of the 100% polystyrene as 89.29 MPa. The strength values of standard polystyrene samples are in the range of 70–100 MPa.

The addition of butyrate kraft lignin to SOPERMA had a large effect on the polymer properties. The T_g 's (as determined from the $\tan \delta$ maximum) of the SOPERMA-styrene polymers as a function of lignin concentration are shown in Fig. 30a. As can be seen, there is a slow increase in T_g at low concentrations and then at 5 wt% lignin the T_g increases significantly.

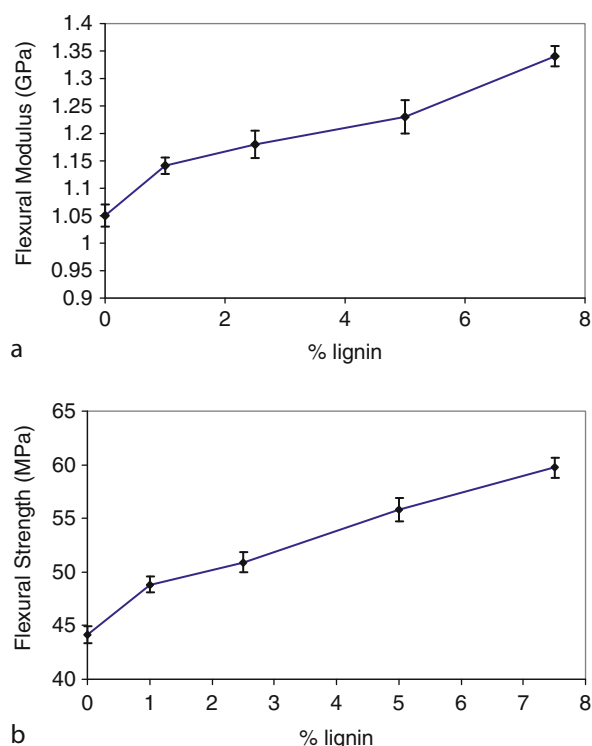
The T_g of the SOPERMA-lignin composite should be influenced by both the cross-link density of the system as well as the higher T_g of kraft lignin (167°C). The cross-link densities, as determined using the modulus in the rubbery region, are shown in Fig. 30b, where the cross-link density of the network increases with lignin until 5 wt% and then starts to decrease again. The increase in cross-link density with lignin addition may be attributed to specific interactions between the polymer matrix and the lignin molecule. The carboxylic acid groups of the SOPERMA monomer may interact with the available hydroxyl groups or thiol groups of the lignin molecule. Additionally, lignin may have effects on the kinetics of polymerization of both the SOPERMA and styrene monomer due to its inhibition effect on radical polymerization, which may affect the cross-link density. However, more work needs to be done to evaluate this effect. The cross-link density levels off at 5 wt% load and starts to decrease for higher concentrations. Thus, at this point the cross-link



New Polymers, Renewables as Raw Materials. Figure 30 (a) Glass transition temperature, T_g , and (b) cross-link density (ν) variation of the SOPERMA-styrene polymers as a function of butyrate lignin content

density must decrease due to the increase in the volume fraction of lignin, which cannot apparently interact with the matrix any more. The significant increase in T_g at 5 and 7.5 wt% lignin shows that the T_g of the system approaches the T_g of the kraft lignin (167°C) component at these high concentrations.

Figure 31 shows that both the flexural modulus and flexural strength of the SOPERMA-styrene polymers increase continuously with increasing lignin content of the resin. Because butyrate kraft lignin is dissolved in the polymer matrix, it should undergo the same strain as the polymer matrix. Thus, the modulus of the composite should increase with the introduction of the rigid aromatic structure of lignin to the system. Additionally, the cross-link density increase with lignin addition is also expected to increase the modulus. On the other hand, the increase in flexural strength with increasing lignin content may be attributed to both the



New Polymers, Renewables as Raw Materials. Figure 31 The change of (a) flexural modulus (b) flexural strength of SOPERMA-styrene polymers at increasing butyrate lignin content

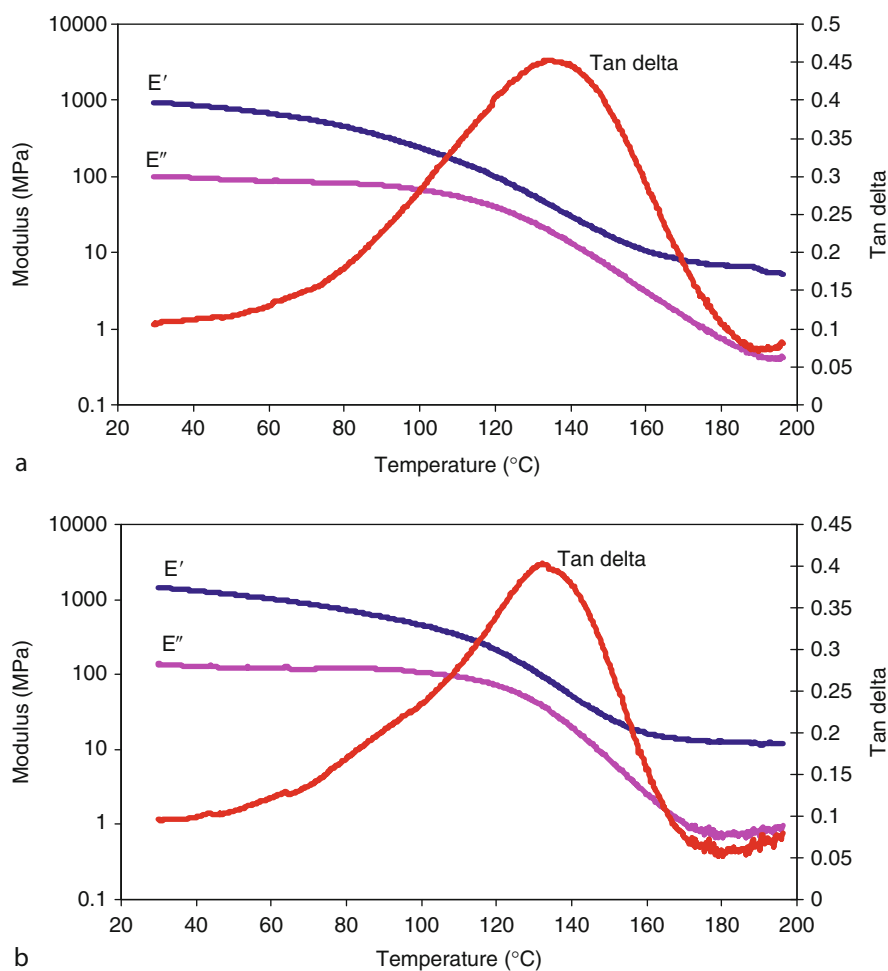
increase in modulus with lignin addition and an increase in cross-link density up to 5 wt% lignin load.

SOGLYME Polymer Properties

As discussed in section “[SOPERMA and COPERMA Polymer Synthesis](#),” the soybean oil monoglyceride methacrylates (SOGLYME) were prepared by methacrylation of the soybean oil glycerolysis product by methacrylic anhydride, as shown earlier in [Fig. 8](#). The methacrylated glycerides did not show phase separation in styrene. The crude methacrylated soybean oil monoglycerides contain methacrylic acid as

a by-product. Methacrylic acid is itself a reactive diluent and acts as a comonomer in the system. Thus, the mechanical properties of the polymers prepared from this crude product (SOGLYME-MEA) as well as the polymers prepared using styrene as the third comonomer (SOGLYME-MEA-ST) will be examined. Properties of the polymers prepared from the acid extracted product with styrene (SOGLYME-ST) will also be examined.

[Figures 32a and b](#) show the storage modulus (E'), loss modulus (E''), and $\tan \delta$ values as a function of temperature for the polymers prepared from the crude methacrylated soybean oil monoglycerides that contain



New Polymers, Renewables as Raw Materials. Figure 32

Storage modulus (E'), loss modulus (E'') and $\tan \delta$ values as a function of temperature for (a) SOGLYME-MEA (b) SOGLYME-MEA-ST polymers

methacrylic acid as a by-product (SOGLEME-MEA) and the styrenated resin (SOGLEME-MEA-ST) (33 wt%) respectively. As can be seen, these polymers also show a broad transition from the glassy state to the rubbery state, similar to the SOPERMA-styrene polymers. The polymers prepared from these resins did not show phase separation, which means that the phase separation observed in the SOPERMA-styrene polymers is not the only factor responsible for the broad transition observed in these triglyceride-based polymers. The plasticization effect of the long flexible fatty acid chains present in the cross-linked monomer has a significant effect on the observed behavior. Additionally, as can be seen, the $\tan \delta$ peak is broader for the SOGLYME-MEA polymer compared to the SOGLYME-MEA-ST polymer showing that the higher cross-link density of the former polymer also has a significant effect on broadening the glass transition.

The E' values, T_g values ($\tan \delta$ maxima) as determined from DMA, flexural moduli, and flexural strengths of the SOGLYME-MEA, SOGLYME-MEA-ST (33 wt% styrene), and SOGLYME-ST (33 wt% styrene) polymers are listed in Table 9. As can be seen, the SOGLYME-MEA-ST polymer has the highest modulus and strength followed by the SOGLYME-MEA polymer. The T_g s of these two polymers do not show a significant difference. The SOGLYME-ST polymer, on the other hand, exhibits considerably lower modulus, strength, and T_g values compared to the other two polymers. The properties of the individual monomers in these polymer systems as well as the cross-link density are both detrimental to the mechanical properties.

New Polymers, Renewables as Raw Materials. Table 9

The 30 °C E' values, T_g values ($\tan \delta$ maxima), flexural modulus and flexural strength of the SOGLYME-MEA, SOGLYME-MEA-ST (33 wt% styrene) and SOGLYME-ST (33 wt% styrene) polymers

Resin	$E'(30^\circ \text{C})$ (GPa)	T_g (°C)	Flexural modulus (GPa)	Flexural strength (MPa)
SOGLEME-MEA	0.86	131.7	0.80	26.5
SOGLEME-MEA-ST	1.15	132.6	1.04	49.0
SOGLEME-ST	0.23	65.5	0.26	1.0

The cross-link densities as determined by using the modulus values in the rubbery region of these polymers are shown in Table 10.

The cross-link densities of the SOGLYME-MEA and SOGLYME-MEA-ST polymers are in the neighborhood of the SOPERMA-styrene polymer at 50 wt% styrene concentration. The mechanical properties observed for these two polymers, however, are much lower than those observed for the SOPERMA-styrene polymers. The SOPERMA-styrene polymer at 50 wt% styrene has a flexural modulus of 1.65 GPa and a flexural strength of 62 MPa. This fact clearly shows that the maleate comonomer compared to the methacrylates, and the styrene comonomer compared to the methacrylic acid, bring more rigidity and strength to the triglyceride-based polymers. The cross-link density of the SOGLYME-MEA-ST polymer is lower than that of the SOGLYME-MEA polymer, as expected due to the increase in comonomer content. The SOGLYME-MEA-ST polymer, despite its lower cross-link density, still shows superior properties compared to those of the SOGLYME-MEA polymer due to the presence of the rigid styrene molecules in the polymer matrix. The significantly lower modulus, strength, and T_g observed for the SOGLYME-ST polymer is unexpected, especially when considering the properties of the SOGLYME-MEA polymer, and can only be attributed to its lower cross-link density compared to the other polymers. It is shown that the fracture strength σ of all bio-based polymers depends on modulus and cross-link density ν , as $\sigma \sim [E\nu]^{1/2}$.

Castor Oil-Based Polymer Properties

The basic fatty acid constituent of castor oil is ricinoleic acid, which is a hydroxy monounsaturated fatty acid

New Polymers, Renewables as Raw Materials. Table 10

The cross-link densities (ν) of the SOGLYME polymers

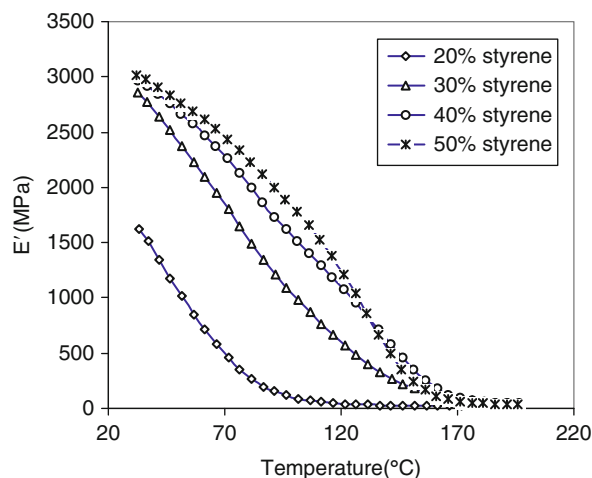
Resin	ν (moles/m ³)
SOGLEME-MEA	701
SOGLEME-MEA-ST	609
SOGLEME-ST	520

(12-hydroxy *cis* 9-octadecenoic acid) (~87%). Castor oil was thus first alcoholized with pentaerythritol, glycerol, and an aromatic diol BPA propoxylate, and the alcoholysis products were then malinated with maleic anhydride, introducing maleate functionality to both the polyol hydroxyls and fatty acid hydroxyls. The resulting resins were labeled COPERMA, COGLYMA, and COBPAPRMA, respectively. The reaction schemes and the structures of the final malinated products for the COGLYMA, COPERMA, and COBPAPRMA products were shown earlier in Figs. 9–11, respectively. Additionally, castor oil was directly malinated and castor oil maleates (COMA) were also prepared. In this section, the properties of these castor oil-based polymers are introduced and analyzed with reference to the network structure. The effect of styrene concentration on mechanical properties of the COPERMA-styrene polymers is also explored and compared to the observed trend for the SOPERMA-styrene polymers.

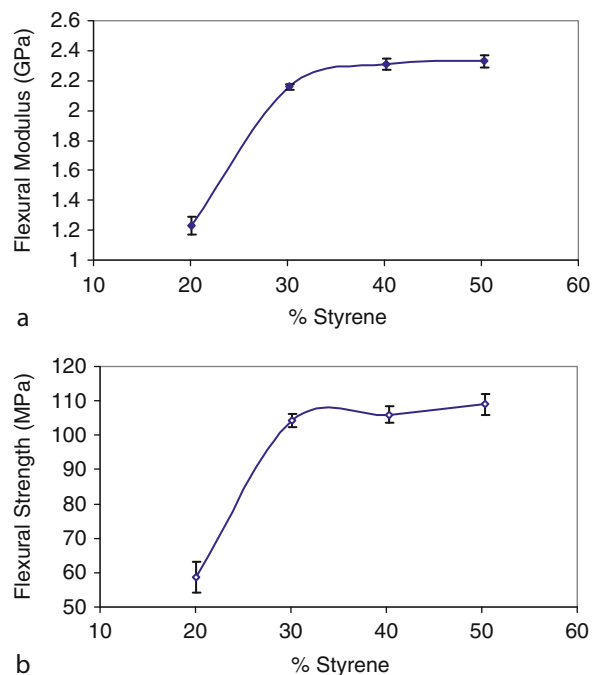
Effect of Styrene Concentration on the COPERMA-Styrene Polymer Properties

Figure 33 shows the storage modulus values, E' , of the COPERMA-styrene polymers as a function of temperature at increasing styrene concentrations as determined from DMA. As can be seen, the room temperature modulus values increase significantly, going from 20 wt% styrene to 30 wt% styrene. The increase in styrene concentration has a much less pronounced effect on the modulus at higher concentrations.

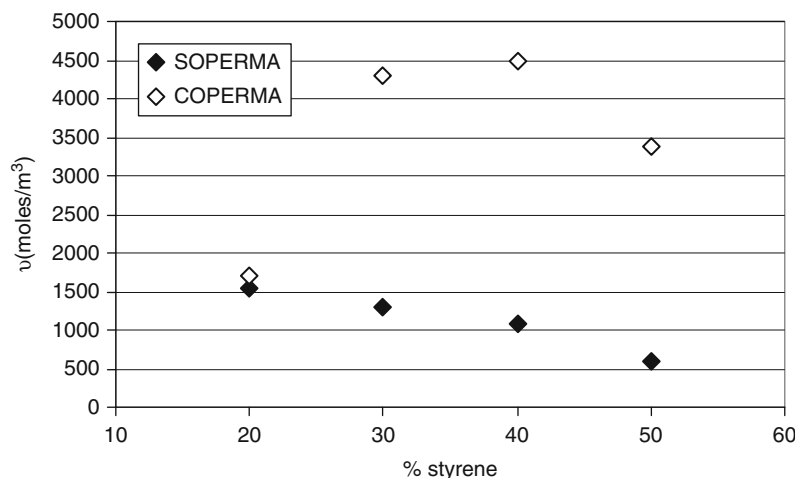
The changes in flexural modulus and flexural strength of the COPERMA-styrene polymers at increasing styrene concentrations are shown in Fig. 34a and b, respectively. The increase in flexural modulus and strength with increasing styrene concentrations follows a trend similar to that observed for the storage modulus. A significant increase in both the modulus and strength is observed while going from 20 to 30 wt% styrene, but this increase levels off rapidly at higher concentrations. As discussed in section “Properties of Plant Oil Resins,” the SOPERMA-styrene polymers showed a continuous increase in both the modulus and strength, with styrene content in similar concentrations.



New Polymers, Renewables as Raw Materials. Figure 33
The change of storage modulus (E') values with temperature for the COPERMA polymers at increasing styrene concentrations



New Polymers, Renewables as Raw Materials. Figure 34
The change in (a) flexural modulus and (b) flexural strength of the COPERMA polymers with increasing concentrations of styrene



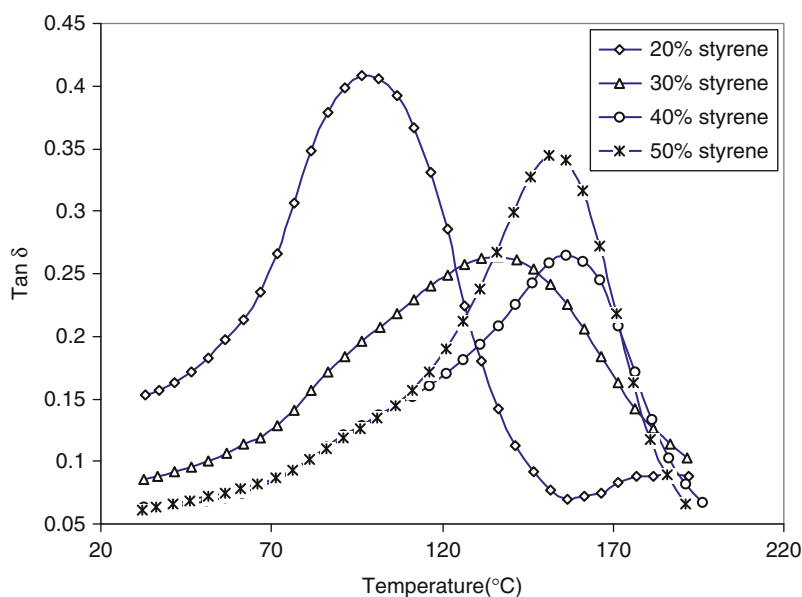
New Polymers, Renewables as Raw Materials. Figure 35

The cross-link densities of the SOPERMA and COPERMA polymers at increasing styrene concentrations

To explain the difference in the effect of styrene concentration on the mechanical properties of these two polymers, it is useful to determine the cross-link densities of the COPERMA-styrene polymers. Figure 35 shows the cross-link densities (ν) of the COPERMA-styrene polymers at increasing styrene concentrations as determined from the rubbery modulus by DMA. The cross-link densities (ν) of the SOPERMA-styrene polymers at the same styrene concentrations are also shown in the same figure, for comparison. For the COPERMA resin, which has a much higher maleate content per triglyceride than the SOPERMA resin, at 20 wt% styrene, the styrene concentration is too low to incorporate all of the maleates into polymerization, since the maleate half-esters do not homopolymerize. The molar ratio of styrene double bonds to maleate double bonds for the 20 wt% styrene COPERMA resin is approximately 0.8, as determined from ¹H-NMR analysis. Thus, the 20 wt% styrene COPERMA polymer has the lowest cross-link density. At 30 and 40 wt% styrene, the molar ratio of styrene double bonds to maleate double bonds is 1.3 and 1.8, respectively, thus a significant increase in cross-link density is observed at 30 wt%, and at 40 wt% the cross-link density reaches its optimum value where all the available reactive groups of the COPERMA monomer can react with styrene. At higher concentrations, the cross-link density starts to decrease again since the added styrene increases the length of the segments between the

cross-links. Thus, the modulus and strength of the COPERMA polymers at 20 wt% styrene are especially low and show a big increase at 30 and 40 wt% styrene due to the significant increase in cross-link density. After this point, the styrene content does not seem to have a significant effect on the modulus and strength. This behavior is different from that of the SOPERMA polymers, which showed a continuous increase in modulus and strength in the same styrene concentrations, despite the decrease in cross-link density. The introduction of the rigid aromatic rings of the styrene comonomer into the SOPERMA monomer with long flexible fatty acid chains results in a higher net increase in both the modulus and strength of the network than that observed for the COPERMA polymers because the fatty acids present in the COPERMA monomer are malinated and therefore incorporated into the network.

The $\tan \delta$ curves for the COPERMA-styrene polymers at increasing styrene concentrations, shown in Fig. 36, also reflect the trend observed in cross-link density. As can be seen, the 20 wt% styrene polymer with the lowest cross-link density shows the $\tan \delta$ maximum at the lowest temperature and therefore has the lowest T_g . The $\tan \delta$ maximum shifts to higher temperatures with increases up to 40 wt% styrene due to the increase in cross-link density. After this point the increase in styrene concentration decreases the cross-link density and the T_g starts to decrease again.



New Polymers, Renewables as Raw Materials. Figure 36

The change of $\tan \delta$ values with temperature for the COPERMA polymers at increasing styrene concentrations

Thus the highest T_g is observed with 40 wt% styrene at 156°C, for the COPERMA-styrene polymers.

Comparison of COPERMA- and SOPERMA-Styrene Polymer Mechanical Properties

As can be seen in Fig. 35, the COPERMA-styrene polymers show significantly higher cross-link densities than those of the SOPERMA-styrene polymers, especially at 30 wt% and higher weight percentages of styrene. Table 11 shows the properties of the 30 wt% styrene SOPERMA and COPERMA polymers for a direct comparison. It can be seen that the modulus value nearly doubles and the flexural strength shows even a larger increase with the change from soybean oil to castor oil in the formulation. The glass transition temperature of the COPERMA polymer is about 9°C higher than that of the SOPERMA polymer. The incorporation of the fatty acid chains into the polymerization both increases the cross-link density and reduces the plasticization effect of the fatty acid chains in the COPERMA polymer, which in turn leads to a considerable increase in modulus, strength, and T_g compared to those properties of the SOPERMA polymer.

For the COPERMA resin, which has the highest maleate content among all the castor oil-based resins,

New Polymers, Renewables as Raw Materials. Table 11

The mechanical properties of COPERMA and SOPERMA polymers at 30 wt% styrene

Property	COPERMA (30% styrene)	SOPERMA (30% styrene)
Flexural strength (MPa)	101.23	43.86
Flexural modulus (GPa)	1.95	1.10
T_g (°C)	144	135
Storage modulus (GPa)	2.88	1.24

the styrene concentration should be more than 30 wt% to fully incorporate all of the maleates during polymerization. This level should be lower for the other resins, which show lower maleate contents than the COPERMA resin. Thus 33 wt% styrene concentration is used for the preparation of other castor oil-based polymers, which keeps the renewable content of the resin within a reasonable range and also gives a chance to compare the properties of these materials

to the commercially available unsaturated polyesters that use similar formulations.

The ratio of the flexural strength of COPERMA/SOPERMA is $\sigma_1/\sigma_2 = 101.2/43.9 = 2.1$. If the square root law for strength is applied, where

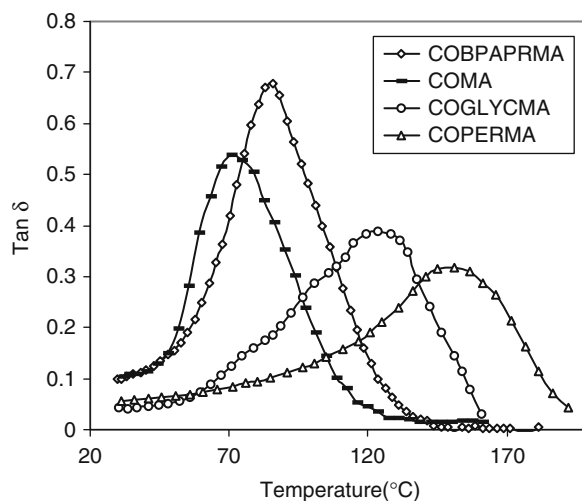
$$\frac{\sigma_1}{\sigma_2} = \left[\frac{E_1 \nu_1}{E_2 \nu_2} \right]^{1/2},$$

then, theory can be readily compared with experiment. Using $E_1 = 1.95$ GPa and $E_2 = 1.1$ GPa (Table 11), $\nu_1 = 4,300$ mol/m³ and $\nu_2 = 1,300$ mol/m³ (Fig. 35 at 30% styrene), then the predicted ratio for strength is $\sigma_1/\sigma_2 = 2.4$, which is in excellent accord with the experimental ratio.

Thermomechanical Properties of Castor Oil-Based Polymers

All of the malinated castor oil-styrene-based polymers exhibited broad transition profiles from the glassy state to the rubbery state, similar to the SOPERMA-styrene polymers. The broad transitions observed were attributed to the phase separation observed on the microscale and the high cross-link density exhibited by these polymers. The plasticization effect of the fatty acids should have a less pronounced effect for the castor oil-based polymers because the hydroxy fatty acids that constitute the majority of the fatty acids (87%) in castor oil were malinated and therefore connected to the network structure. Figure 37 shows the $\tan \delta$ curves of the castor oil-based polymers as determined by DMA.

As discussed above, the damping peak position is sensitive indicator of cross-linking. As the cross-link density increases, the $\tan \delta$ maximum shifts to higher temperatures, the peak broadens and a decrease in the $\tan \delta$ value is observed. The cross-link densities of the malinated castor oil-based polymers (33 wt% styrene) determined by using the value of E in the rubbery region of the polymers as determined by DMA are shown in Table 12. As can be seen, the COPERMA polymer has the highest cross-link density, followed by COGLYMA, COMA, and COBPAPRMA polymers. The COBPAPRMA polymer shows the lowest cross-link density, despite the higher maleate content than both COMA and COGLYMA which can be attributed to the bulkiness of the Bisphenol



New Polymers, Renewables as Raw Materials. Figure 37
The $\tan \delta$ values of the castor oil polymers as a function of temperature

New Polymers, Renewables as Raw Materials. Table 12
Crosslink densities of castor oil polymers

Resin type	Mc(g/mole)	ν (moles/m ³)
COPERMA	255	4,310
COGLYMA	441	2,494
COMA	732	1,511
COBPAPRMA	783	1,418

A propoxylate moiety. As can be seen in Fig. 37, the COPERMA polymer with the highest cross-link density exhibits the highest T_g and shows the broadest peak, with the lowest $\tan \delta$ value. The COBPAPRMA, which has the lowest cross-link density, shows the highest $\tan \delta$ values as expected; however, its $\tan \delta$ max temperature is about 14° C above than that of the COMA polymer. The higher T_g observed for the COBPAPRMA may be explained by the presence of the rigid aromatic backbone of BPA propoxylate, as compared to the aliphatic fatty acid backbone of the COMA polymer. Thus, the monomer chemical structural influence dominates the cross-link density effect on T_g .

Mechanical Properties of Castor Oil-Based Polymers

The storage modulus values at 30°C and the T_g 's as determined from DMA, as well as the flexural modulus, flexural strength, and the surface hardness values of the castor oil polymers are given in Table 13. The styrene content of each resin was 33 wt%. The mechanical property hardness is the ability of the material to resist indentation, scratching, abrasion, cutting, and penetration. This property may be important for structural materials that require a high resistance to indentation or abrasion. The hardness of a polymer reflects such other qualities as resilience, durability, uniformity, strength, and abrasion resistance. As can be seen in Table 13, the surface hardness of the castor oil-based polymers changes proportionately with the strength of the polymers.

The observed mechanical properties of the castor oil-based polymers can be explained both in terms of the cross-link density and the chemical structures of the polyols used. The COPERMA polymer, which has the highest cross-link density, shows superior properties to the other castor oil-based polymers. The COPERMA polymer with its T_g around 150°C and flexural modulus of 2.2 GPa and flexural strength of 105 MPa exhibits the highest T_g and strength obtained from any triglyceride-based thermoset resins. The COBPAPRMA polymer's modulus, strength and surface hardness values are higher than those of both COMA and COGLYMA polymers and approach those of COPERMA polymer, although its cross-link density is slightly lower than these two polymers. The aromatic structure of the BPA propoxylate moiety brings both rigidity and strength to the polymer network. Thus,

this resin shows both high modulus and strength with a reasonable T_g , despite its lower maleate content, which is beneficial for the formulation since it decreases the nonrenewable content of the polymer. The COMA polymer shows the lowest modulus, strength, surface hardness, and T_g values due to its low cross-link density and also shows that a multifunctional unit at the center of the triglyceride monomer structure is essential for improved properties for these polymers.

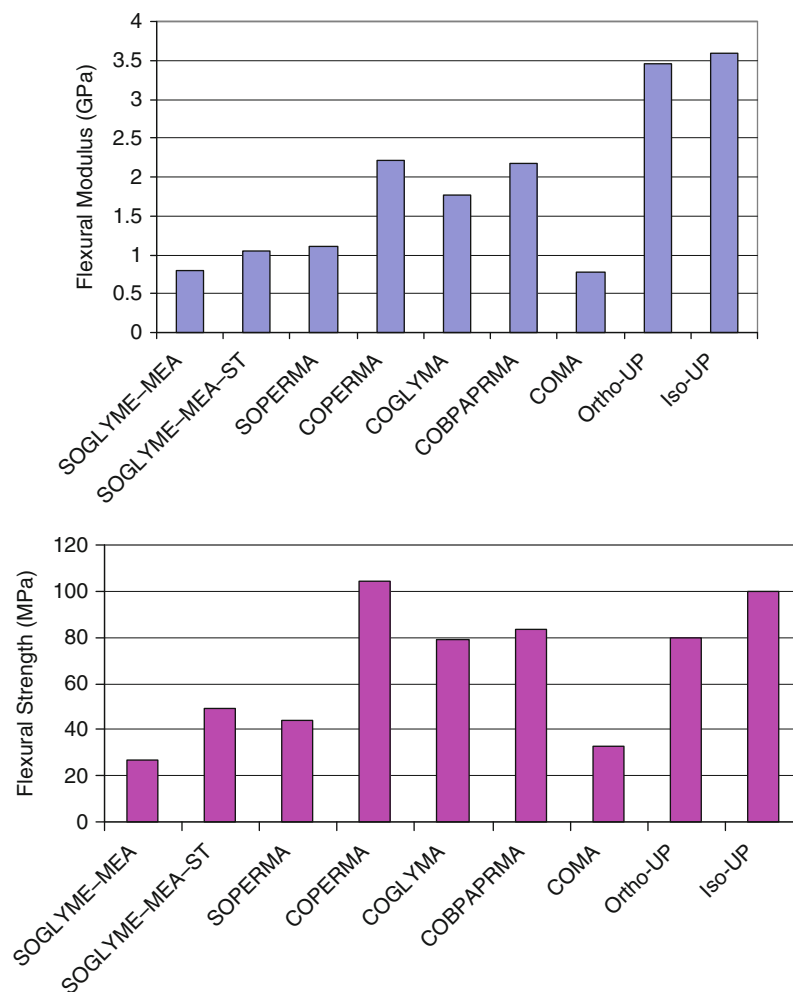
Figure 38 gives a comparison between the mechanical properties of these bio-based resins and the properties of two of the most commonly used UP resins: Orthophthalic (Ortho-UP) and Isophthalic (Iso-UP) UP resins. As can be seen, the properties of castor oil-based polymers are in a comparable range with those of the commercially successful UP resins. The properties of these bio-based resins can be significantly improved by the addition of lignin, which introduces the required aromatic groups for high stiffness and high T_g polymers and further increases their bio-based content with low-cost renewable material.

Future Directions

Triglyceride oils are an abundant natural resource that has yet to be fully exploited as a source for polymers and composites. The different chemical functionalities allow the triglyceride to be converted to several promising monomers. These are being commercialized for applications in composite resins, foams, pressure sensitive adhesives, elastomers, coatings with significantly reduced toxicity, and carbon footprint compared to their petroleum-based counterparts. When blended

New Polymers, Renewables as Raw Materials. Table 13 The mechanical properties of castor oil polymers

Resin type	E' (30°C) (GPa)	T_g (°C)	Flexural modulus (GPa)	Flexural strength (MPa)	Surface hardness (D)
COPERMA (1:2:10.7)	2.94	150	2.21	101.60	89.3
COGLYMA (1:2.2:9.2)	2.40	124	1.76	78.89	86.1
COBPAPRMA (1:2:6.7)	2.69	86	2.17	83.20	88.5
COMA (1:3)	1.15	72	0.78	32.83	77.1
Ortho-UP	–	135–204	3.45	80	–
Iso-UP	–	135–204	3.59	100	–



New Polymers, Renewables as Raw Materials. Figure 38

A property comparison of soyoil and castor oil based polymers with commercial petroleum based Ortho and Iso-Unstaturated Polyester resins

with comonomers, these monomers form polymers with a wide range of physical properties. They exhibit flexural strength up to 100 MPa and tensile moduli in the 1–3 GPa range, with T_g ranging from 70°C to 150°C, depending on the particular monomer and the resin composition. DMA shows that the transition from glassy to rubbery behavior is extremely broad for these polymers as a result of the triglyceride molecules acting both as cross-linkers and as plasticizers in the system. Saturated fatty acid chains are unable to attach to the polymer network, causing relaxations to occur in the network. However, this transition can be sharpened by reducing the saturation content, as

demonstrated with the genetically engineered oil and pure triolein oil. The new design rules for controlling the molecular architecture based on the Twinkling Fractal Theory of the Glass Transition [67–69] will allow the facile development of these materials in future years. When combined with natural fibers (chicken feathers, hemp, flax, straw, etc.), useful high volume applications are made possible in hurricane resistant housing, positive net energy housing, earthquake resistant structures, eco-leather, and wind turbines in support of renewable energy.

This area of research sets a foundation from which completely new materials can be produced with novel

properties. Work continues on optimizing the properties of these green materials and understanding the fundamental issues that affect them. Computer simulation can be used to optimize the choice of the fatty acid distribution function and determine the resulting architecture and mechanical properties for the particular chemical pathways shown in Fig. 2. Use of a computer significantly reduces the number of chemical trials required for a system with a large number of degrees of freedom and suggests the optimal oil most suited to a particular type of resin. In this manner, more renewable resources can be used to meet the material demands of many industries.

Bibliography

- Wool RP, Kusefoglu S, Zhao R, Palmese G, Khot S (2000) High Modulus Polymers and Composites from Plant Oils. U.S. Patent 6,121,398
- Wool RP (1999) *Chemtech* 29:41
- Liu K (1997) Soybeans: chemistry, technology, and utilization. Chapman and Hall, New York
- Gunstone F (1996) Fatty acid and lipid chemistry. Blackie Academic and Professional, New York
- Cunningham A, Yapp A (1971) U.S. Patent 3,827,993
- Bussell GW (1971) U.S. Patent 3,855,163
- Hodakowski LE, Osborn CL, Harris EB (1975) U.S. Patent 4,119,640
- Trecker DJ, Borden GW, Smith OW (1976) U.S. Patent 3,979,270
- Trecker DJ, Borden GW, Smith OW (1976) U.S. Patent 3,931,075
- Salunkhe DK, Chavan JK, Adsule RN, Kadam SS (1992) World oilseeds: chemistry, technology, and utilization. Van Nostrand Reinhold, New York
- Force CG, Starr FS (1988) U.S. Patent 4,740,367
- Barrett LW, Sperling LH, Murphy CJ (1993) *J Am Oil Chem Soc* 70:523
- Qureshi S, Manson JA, Sperling LH, Murphy CJ (1983) In: Carraher CE, Sperling LH (eds) *Polymer applications of renewable-resource materials*. Plenum Press, New York
- Devia N, Manson JA, Sperling LH, Conde A (1979) *Polym Eng Sci* 19:878
- Devia N, Manson JA, Sperling LH, Conde A (1979) *Polym Eng Sci* 19:869
- Devia N, Manson JA, Sperling LH, Conde A (1979) *Macromolecules* 12:360
- Sperling LH, Carraher CE, Qureshi SP et al (1991) In: Gebelein CG (ed) *Polymers from biotechnology*. Plenum Press, New York
- Sperling LH, Manson JA, Linne MA (1984) *J Polym Mater* 1:51
- Sperling LH, Manson JA (1983) *J Am Oil Chem Soc* 60:1887
- Fernandez AM, Murphy CJ, DeCosta MT et al (1983) In: Carraher CE, Sperling LH (eds) *Polymer applications of renewable-resource materials*. Plenum Press, New York
- Sperling LH, Manson JA, Qureshi SA, Fernandez AM (1981) *Ind Eng Chem* 20:163
- Yenwo GM, Manson JA, Pulido J et al (1977) *J Appl Polym Sci* 21:1531
- Frischinger I, Dirlikov S (1991) *Polymer Comm* 32:536
- Frischinger I, Dirlikov S (1994) In: Sperling LH, Kempner D, Utracki L (eds) *Interpenetrating polymer networks. Advances in chemistry series 239*. American Chemical Society, Washington, DC, p 517
- Rosch J, Mulhaupt R (1993) *Polymer Bull* 31:679
- Meffert A, Kluth H (1989) Denmark Patent 4,886,893
- Rangarajan B, Havey A, Grulke EA, Culnan PD (1995) *J Am Oil Chem Soc* 72:1161
- Zaher FA, El-Malla MH, El-Hefnawy MM (1989) *J Am Oil Chem Soc* 66:698
- Friedman A, Polovsky SB, Pavlichko JP, Moral LS (1996) U.S. Patent 5,576,027
- Swern D, Billen GN, Findley TW, Scanlan JT (1945) *J Am Oil Chem Soc* 67:786
- Sonntag NOV (1982) *J Am Oil Chem Soc* 59:795
- Solomon DH (1967) *The chemistry of organic film formers*. Wiley, New York
- Can E (1999) M.S. Thesis, Bogazici University, Turkey
- Bailey AE (1985) In: Swern D (ed) *Bailey's industrial oil and fat products*. Wiley, New York
- Hellsten M, Harwigsson I, Brink C (1999) U.S. Patent 5,911,236
- Cain FW, Kuin AJ, Cynthia PA, Quinlan PT (1995) U.S. Patent 5,912,042
- Eckwert K, Jeromin L, Meffert A, et al (1987) U.S. Patent 4,647,678
- Khot SN (1998) M.S. Thesis, University of Delaware
- Wypych J (1986) *Polyvinyl chloride stabilization*. Elsevier, Amsterdam
- Sears JK, Darby JR (1982) *The technology of plasticizers*. Wiley, New York
- Carlson KD, Chang SP (1985) *J Am Oil Chem Soc* 62:931
- Raghavachar R, Letasi RJ, Kola PV et al (1999) *J Am Oil Chem Soc* 76:511
- Pashley RM, Senden TJ, Morris RA, et al (1991) U.S. Patent 5,360,880
- Likavec WR, Bradley CR (1999) U.S. Patent 5,866,628
- Bordon GW, Smith OW, Trecker DJ (1971) U.S. Patent 4,025,477
- La Scala JJ, Wool RP (2002) *J Am Oil Chem Soc* 79:59
- Bunker SP (2000) M.S. Thesis, University of Delaware
- Chu TJ, Niou DYJ (1989) *Chin Inst Chem Eng* 20:1
- Betts AT (1975) U.S. Patent 3,867,354
- Mitch EL, Kaplan SL (1975) In: *Proceedings of 33 rd annual SPE technical conference*, Atlanta
- Can E, Kusefoglu S, Wool RP (2001) Rigid thermosetting liquid molding resins from renewable resources: I. Copolymers of soy oil monoglycerides with maleic anhydride. *J Appl Polym Sci* 81:69
- Gardner HC, Cotter RJ (1981) European Patent 20,945
- Thomas P, Mayer J (1971) U.S. Patent 3,784,586
- Lee SH, Park TW, Lee SO (1999) *Polymer (Korea)* 23:493
- Shioneh, Yamada J (1999) Japanese Patent 11,147,222

56. Hasegawa H (1999) Japanese Patent 11,240,014
57. Johnson LK, Sade WT (1993) *J Coat Tech* 65:19
58. Solomons TWG (1992) *Organic chemistry*. Wiley, New York
59. La Scala J, Wool RP Polymer. in preparation
60. Flory PJ (1975) *Principles of polymer chemistry*. Cornell University, Ithaca
61. Nielsen LE, Landel RF (1991) *Mechanical properties of polymers and composites*. Marcel Dekker, New York
62. Wool RP, Khot SN (2000) In: *Proceedings of ACUN-2, Sydney*
63. Can E, Kusefoglu S, Wool RP (2001) Rigid thermosetting liquid molding resins from renewable resources: (I) copolymers of soy oil monoglycerides with maleic anhydride. *J Appl Polym Sci* 81:69
64. Khot SN, La Scala JJ, Can E et al (2001) *J Appl Polym Sci* 82:703
65. Can E, Kusefoglu S, Wool RP (2002) Rigid thermosetting liquid molding resins from renewable resources. II. Copolymers of soybean oil monoglyceride maleates with neopentyl glycol and bisphenol A maleates. *J Appl Polym Sci* 83(5):972–980
66. Can E (2004) PhD Thesis, University of Delaware
67. Wool RP (2008) *J Polym Sci Part B: Polym Phys* 46:2765
68. Wool RP, Campanella A (2009) *J Polym Sci Part B: Polym Phys* 47:2578
69. Stanzione JF III, Strawhecker KE, Wool RP, doi:10.1016/j.jnoncrysol.2010.06.041

Nickel-Based Battery Systems

RALPH J. BRODD

Broddarp of Nevada, Henderson, NV, USA

Article Outline

Glossary

Definition of the Subject

Introduction

Characteristics of the Nickel Active Materials

Nickel-Iron Batteries (Ni-Fe)

Nickel Cadmium Batteries

Nickel Metal Hydride

Nickel-Zinc Batteries

Nickel-Hydrogen Batteries

Bibliography

Glossary

β - β Transformation The reaction of β -Ni(OH)₂ on charge to form β -NiOOH.

α - γ Transformation The reaction of α -Ni(OH)₂ on charge to form γ -NiOOH.

Ah Ampere hour 3,600 coulombs, the quantity of current flow in 1 ampere in 1 h.

Charge reserve Additional capacity of the negative or positive electrode to prevent gas evolution when the cell is overcharged or overdischarged.

Electrolyte Electrical conducting liquid flow where charge is carried by positive and negative ions.

Flame-arresting vent A cell vent designed to stop burning discharge from a vent.

Ni-MH Nickel metal hydride cell containing a nickel hydroxide positive and a hydrogen-absorbing metal alloy negative electrode.

Plate Electrode construction containing the active material and a current collector.

Pocket plate electrodes Perforated nickel-plated steel pockets that contain the active materials in a cell with free access of electrolyte.

Self-discharge Loss of capacity due to the chemical instability or a reaction of an active material in an electrode with the electrolyte.

Tubular plate electrodes Perforated nickel-plated steel tubes used to contain the active material in a cell.

Misch metal Containing a mixture of hydrogen-absorbing transition metals, Mm.

Nickel cadmium fiber plate Batteries construction of the positive nickel battery plate that contains nickel metal fibers throughout the plate for more efficient current collection.

Sinter plate cells Sealed battery cells using a nickel current collector structure usually produced by heating to a temperature where powdered nickel metal particles bond together to form a porous structure sinter current collector.

Trickle charging The low-level current flow to maintain full charge in a battery.

Terminal The external connection to the positive and negative electrodes containing the active material inside the battery case.

WH Watt hour, energy content of a battery expressed as the product of ampere-hours times cell voltage.

Zircar Trademark of nonwoven ceramic separator used in some Ni-MH cells.

Definition of the Subject

Nickel batteries are rechargeable batteries that are used in a variety of applications including portable electronic

devices, electric and hybrid vehicles, aeronautics and aerospace and stationary energy storage among others. They operate over a wide temperature range, have a flat discharge curve and are available in sizes ranging from small coin cells to motive power batteries. Nickel batteries are physically and electrically rugged and abuse tolerant including over charge and overdischarge.

Introduction

The nickel electrode serves as cathode for several important commercial rechargeable battery systems. The characteristics of these systems are listed in Table 1. The first commercial nickel battery was the nickel-iron system which provided lighting in railroad cars due to its strong resistance to physical and electrical abuse. The electrode structure has a strong influence on the operating life of a battery system. The nickel systems are robust, both physically and chemically.

Nickel battery systems compete directly with the lead acid battery in many commercial energy storage applications and with Li-Ion in portable electronic applications.

Advantages	Disadvantages
Long cycle life	Heavier than Li-Ion
High reliability	Higher cost than lead acid
Excellent rate capability	
Very robust cell constructions	

Characteristics of the Nickel Active Materials

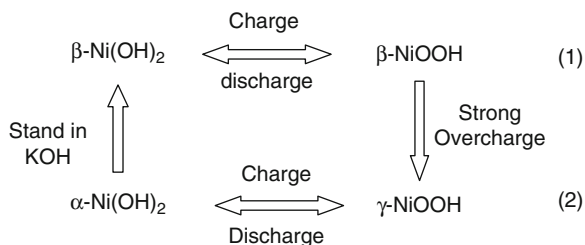
The family of nickel batteries is based on the utility, strength, and reversibility of the nickel electrode reactions in alkaline media. The nickel active materials for use in batteries are produced, mainly, by chemical precipitation of $\text{Ni}(\text{OH})_2$ with the addition of KOH to aqueous nickel sulfate solutions made by dissolving nickel metal in sulfuric acid. The nickel active material is usually produced by a precipitation process wherein a nickel sulfate solution is heated and pumped into hot NaOH. The process produces $\beta\text{-Ni}(\text{OH})_2$. Fine graphite may be added, depending on the application. A small percentage of cobalt, ranging from 3% to 7%, depending on application, is co-precipitated with the nickel sulfate solution to improve charge acceptance, capacity, and operating life. The precipitate is filtered, dried, and ground. Care must be exercised to keep the drying temperature below 200°C as higher temperatures cause formation of electrochemically inactive bunsenite structure. Electrodes for cells used in satellites and other space applications are made by an in situ electrochemical precipitation technique described later.

Nickel Electrode Reaction Characteristics

The battery electrolyte is aqueous KOH solution in a concentration range of 25–30% KOH by weight. The mechanism that best describes the overall nickel electrode reactions in alkaline electrolyte is given in Eqs. 1 and 2 below.

Nickel-Based Battery Systems. Table 1 Characteristics of Nickel batteries (Nominal values)

System	Cell voltage	Wh/kg	Wh/l	Cycle life, 100% DOD	Calendar life, years	Commercial introduction
Ni-Fe	1.37	30	55	2,000–4,000	25	1,898
Ni-Cd, Vented pocket plate	1.29	20	40	500–2,000	25	1,900
Ni-Cd, Vented sintered plate	1.29	30	58	500–2,000	10	1,940
Ni-Cd, Sealed sinter plate	1.29	35	100	300–700	5	1,950
Ni-Cd, Fiber	1.29	10	15	500–10,000	20	1,950
Ni-MH, Sealed sinter plate	1.2	110	430	300–1,000	5	1,980
Ni- H_2	1.4	60	65	1,500–10,000	25+	1,980
Ni-Zn	1.71	60	120	200–500	3–5	1,931, 2,009



Equation 1 describes the classic β - β transformation reaction of the nickel electrode in alkaline media wherein the valence of the nickel shuttles between divalent Ni^{+2} and trivalent Ni^{+3} . The charged and discharged materials have similar structures and long cycle life. The battery reactions follow this regimen with little or no overcharge. On continued charging after reaching the Ni^{+3} valence state, the high oxygen overvoltage on the electrode surface permits formation of a Ni^{+4} valent nickel in the electrode structure. This transformation stops with oxygen evolution on the electrode surface and a nickel valence of 3.67, resulting in a higher capacity cell. On discharge the γ -NiOOH returns to α -Ni(OH)₂. Charge-discharge will continue on this higher capacity regime. However, if the cell is left on open circuit, the Ni(OH)₂ slowly transforms back to β -Ni(OH)₂. Transition back to the β - β cycle occurs only on open circuit stand as the α structure transforms slowly back to the β structure. The normal materials in electrode construction are produced by chemical or electrochemical precipitation, depending on the application for the cell. The uncharged β -Ni(OH)₂ is usually used in electrode fabrication and cells are initially assembled in the discharged state.

The four nickel materials in Eqs. 1 and 2 are distinct phases and have been characterized both-electrochemically and by the XRD patterns exhibited in the chemically precipitated phases. A point defect, nonstoichiometric, structural model correctly describes the structure and interactions of nickel electrode active materials (22). In this model, the intercalated water is incorporated as interlamellar protons, with the oxygen effectively adding to the NiO₂ layers. The model introduces nickel vacancies that explain the empirical Ni/O ratios (less than 1/2). The Raman spectra gives no indication of water in the lattice as the excess protons and alkali cations occupy nickel vacancies or interlamellar sites and three protons can occupy a nickel site for charge compensation. This model

correctly explained the empirical observation that Ni has a 3.67 oxidation state observed after overcharge. Raman spectra of these higher oxidation state materials are consistent with the simple rhombohedral space group and the simple γ -NiOOH-type crystal lattice (the ideal, stoichiometric crystal). The spectra are inconsistent with the larger space groups and those which contain molecular water. Thus, the structure is written ideally as γ -NiOOH.

There have been four generations of nickel electrode construction. The first generation included the pocket and tubular electrode structures that were developed independently by Junger in Sweden and Edison in the USA in 1900 for the original system introduction. The Sintered Plate is a second-generation technology and was developed in the 1940s by Varta. The Nickel Cadmium Fiber electrode battery, developed by DAUG, constitutes the third-generation (1985) technology. The nickel cathode electrodes used in nickel-hydrogen batteries for space applications constitute the fourth generation and are produced by an electrochemical deposition of the nickel hydroxide materials directly into the voids in the sintered nickel electrode structure. This is a more expensive process and is used mainly for space and special applications. Because of their ability to withstand physical and electrical abuse, nickel batteries are virtually indestructible. All four electrode constructions are articles of commerce in 2011. Their characteristics will be discussed in more detail below.

Nickel-Iron Batteries (Ni-Fe)

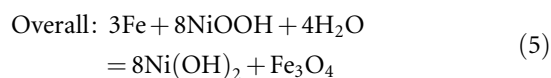
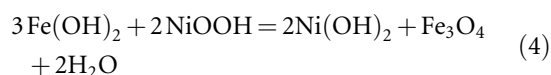
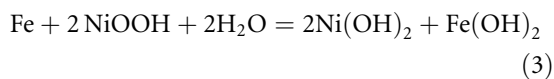
Pocket and Tubular Plate Cells

The iron-nickel battery system is one of the oldest rechargeable systems. It was invented about the turn of the century, independently in Sweden by Junger in 1899, and in the USA by Edison in 1900, each acting without specific knowledge of the other's activity. Common usage today is to call the system the Edison battery. It still finds application today to power materials handling vehicles, underground mining vehicles, miners' lamps, railway cars and signal systems, and emergency lighting. The Edison battery is virtually indestructible and can withstand both physical and electrical abuse situations, discharged stand, and short-circuiting. Edison used tubes for the positive

and pockets for the negative electrode structures for his 1908 battery designs. Disadvantages include low power density, low temperature performance, poor charge retention, and hydrogen evolution on stand.

Both the positive and negative tubular and pocket current collectors are made of perforated nickel-plated steel. They are very robust and are virtually indestructible. The low energy density, poor charge retention, and poor low temperature performance, along with high cost of manufacture, have led to a decline in use of the nickel-iron battery system. The negative electrode, or anode, is iron and the positive electrode, or cathode, is nickel oxide with 6–8 molar potassium hydroxide (KOH) as the electrolyte. The cell reaction occurs in a two-step process. The system has a nominal voltage of 1.37 V.

The two-step cell reaction is:



On discharge, iron initially forms a divalent hydroxide Fe(OH)_2 in the first plateau. Further discharge forms Fe(OH)_3 , which then further reacts with Fe^{+2} to form Fe_3O_4 .

Advantages	Disadvantages
Physically robust	High self-discharge
Long stand and cycle life	Hydrogen evolution on charge and discharge
Withstand electrical abuse	Low voltage and energy density
Very reliable	Poor low temperature performance

The construction of the tubular and pocket plate nickel-iron battery cell is essentially identical to that of the nickel cadmium battery and has not changed over the past 50 years. For good performance, special attention must be paid to use high purity materials and the particle size characteristics of the active materials. The

long cycle life is a result of the low solubility of the reaction intermediates. On discharge, supersaturation of the discharge product results in the formation of minute crystallites of Fe_3O_4 near the reaction site. On recharge, these crystals form into a high surface area iron with excellent high rate performance on discharge.

Preparation of Negative Iron Electrode Materials

The active material is produced by dissolution of pure iron in sulfuric acid. The FeSO_4 is recrystallized, dried, and roasted at high temperature to produce Fe_2O_3 . The materials are washed free of the sulfate, dried, and partially reduced in H_2 to produce a mixture of Fe_3O_4 and iron. The resulting mixture is blended with small amounts of FeS, sulfur, and HgO for use in negative plate assembly.

The anode current collector is an annealed, perforated, nickel-plated steel pocket plate assembly. The tubes for the positive electrode are produced from perforated nickel-plated strips that are wound to produce a tube. One end is crimped and the active material poured into the open end in layers and crimped again to close the tube. A machine automatically introduces the active material and tamps it into the pockets. After filling, the tubes are pressed into openings in the nickel steel electrode frame. The pocket plate follows a similar process. Rectangular pockets of perforated nickel-plated iron strips are filled with active material, crimped, closed, and fixed/bolted in a nickel-plated steel frame. The assembled cells are placed in polyethylene containers and filled with KOH electrolyte. Spacing of the negative and positive electrodes is maintained by the internal assembly structure. There are no separators as are common in the lead acid battery structure.

Preparation of Positive Nickel Electrode Materials

The preparation of the Ni(OH)_2 active material starts with dissolving a high purity nickel metal powder, or chips, in sulfuric acid. The hydrogen produced in this step is used in making the negative iron active material. The acidity is adjusted to pH 3 or 4 to remove iron and other insoluble materials. Further conditioning, to remove all of the ferrous iron and copper, may be needed. Cobalt sulfate is added at about the 1.5% level to enhance the charge acceptance of the nickel electrode performance. The resulting solution is heated

and sprayed into hot 50% NaOH to precipitate the $\text{Ni}(\text{OH})_2$ active material. The resultant is dried, crushed, and screened to pass 20- but not 200-mesh screen. The construction of the negative plate follows a similar process to that of the positive plate described above. The size of the tubes and configuration of the plates determine the capacity of the finished battery. The battery capacity is designed to be limited by the capacity of the positive plate for longer cycle life.

The nickel-iron battery cell fabrication process is essentially unchanged in over 50 years. Special attention must be paid to use high purity materials and particle size characteristics of the active materials. The iron negative active material is made from pure iron that is dissolved in sulfuric acid. The resulting $\text{Fe}(\text{SO}_4)_2$ is recrystallized and dried. This is washed free of sulfuric acid and roasted at 915°C to form a mixture of Fe_2O_3 and Fe metal and is, then, blended with small amounts of FeS, sulfur, and HgO for use in the negative plate assembly.

The negative current collector is a perforated nickel-plated steel pocket plate assembly. The tubes are produced from perforated nickel-plated strips that are wound to produce a tube. The tubular construction has one end crimped and the active material is poured into the open end in layers with nickel flakes. Alternatively, a machine automatically introduces the active material and tamps it into the pockets. After filling, the tubes/pockets are crimped and pressed into openings in the nickel-plated steel electrode frame. A similar process is used for the positive electrode.

The positive electrode consists of $\text{Ni}(\text{OH})_2$ in alternate layers of nickel flake. High purity nickel metal powder or chip is dissolved in sulfuric acid. Hydrogen is used in making the negative iron active material. The acidity is adjusted to pH 3 or 4 to remove iron and other insoluble materials. Further conditioning to remove all of the ferrous iron and copper may be needed. Cobalt sulfate is added at about the 1.5% level to improve the nickel performance. The resulting sulfate solution is heated and sprayed into hot 50% NaOH to precipitate the $\text{Ni}(\text{OH})_2$ active material. The resulting slurry is filtered, washed, dried, and screened to control the particle size to pass 20 but not 200 mesh.

The preferred electrolyte is 25–30% KOH solution with up to 50 g/l of LiOH added. The addition of LiOH improves cell capacity, prolongs cycle life,

and facilitates the kinetics of the nickel electrode. It expands the discharge plateau and delays oxygen evolution. It also decreases the carbonate content of the electrolyte as Li_2CO_3 is not very soluble. The negative and positive plates are usually separated by hard rubber pins. The edges of the plates are insulated from the sides and bottom by hard rubber or polymer frames. The spring-loaded, hinged, filler cap protects against carbonation of the alkaline electrolyte by atmospheric CO_2 .

The nickel flake is produced by the dissolution of high purity nickel powder in sulfuric acid and the H_2 used in the production of the negative iron active material. The acidity is adjusted to remove soluble iron and copper. A small amount of cobalt sulfate may be added to improve performance. The resulting solution is sprayed into a hot concentrated NaOH solution. The construction of the positive plate follows that for the negative plate.

The size of the tubes and configuration of the plates determine the capacity of the finished battery. The positive and negative electrode frames are bolted together with a spacer to separate the negative from the positive frames for strength and long life. The battery is designed to be limited by the capacity of the positive electrode for longer cycle life. Constant potential charging is not recommended as the system has a tendency to go into thermal runaway. A modified constant voltage with a current limit is acceptable.

Performance

The Ni-Fe batteries range in size from 5 Ah to over 1,200 Ah. The cell open circuit voltage of the Ni-Fe cell is 1.4 which drops quickly to 1.2 V during discharge. Tubular or pocket plate constructions are used. Active materials of high purity are contained inside the perforated nickel-plated steel tubes or rectangular pockets. The active materials are iron for the negative electrode and nickel oxide for the cathode and a KOH electrolyte.

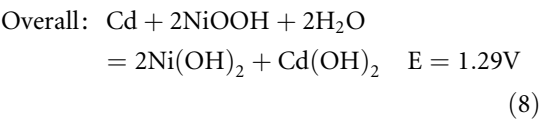
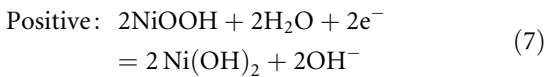
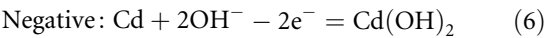
The battery is assembled in the discharged state and a formation charge is used to activate it. The KOH electrolyte is invariant over the charge-discharge regime. The open circuit voltage is 1.4 V that drops quickly to 1.2 V on discharge. The cells vent hydrogen on charge and require a well-ventilated space to prevent

accumulation of hydrogen that could ignite with a spark. The low solubility of iron on discharge results in the formation of minute crystallites of Fe_3O_4 near the reaction site. On charge, these crystallites result in the formation of a high surface area iron electrode deposition that increases the ability to deliver pulse and high current discharge.

Constant potential charging is not recommended as the system has a tendency to go into thermal runaway. A modified constant voltage with a current limit is preferred to prevent thermal runaway situations. The long cycle life of the Ni-Fe battery is a result of the low solubility of the iron electrode reaction intermediates. Series-connected cells require precautions in dealing with high voltage cell assembly connections. Batteries that will be inactive for a month or more should be discharged, short-circuited, and left in that condition. They can be stored in the discharged state with filling caps closed for long periods of time without damage.

Nickel Cadmium Batteries

Nickel Cadmium batteries are available in four different constructions: vented pocket plate, vented tubular plate, sealed sinter plate, and fiber plate constructions. The vented pocket and tubular electrode constructions followed directly from the nickel-iron batteries of Junger and Edison with a substitution of cadmium for iron in the electrode. Like the nickel-iron system, the nickel cadmium pocket and tubular systems are very sturdy, have long life and operate over a wide temperature range with good high rate performance, and require little maintenance. Because they can withstand overcharge, overdischarge, mechanical and electrical abuse, and require little maintenance, they find use in critical industrial applications including switchgear operations, telecommunications, aircraft power, uninterruptible power supplies (UPS), and emergency lighting even though their cost is higher than the lead acid battery system. The cell voltage is 1.29 V and the cell electrolyte is KOH with added LiOH to improve cycle life and high temperature operation.



Advantages	Disadvantages
Very rugged	Low energy density
Long cycle life	Higher cost than lead acid
Reliable	Contains cadmium
Flat discharge profile	Oxidation of carbon conductor forms CO_2
	Vented cells need regular water service

Nickel cadmium batteries have long life as measured by the number of charge-discharge cycles a battery can sustain and still deliver useful capacity. Under normal conditions a typical Ni-Cd cell can deliver 2,000 cycles or more over a period of 5–20 years depending on application (drain rate, temperature, and ambient conditions of temperature, vibration, duration of high current demand and recharging conditions) as well as the internal design of the cell. Typically, the Ni-Cd cell has a strong internal mechanical design, chemically stable internal ambient and ability to withstand electrical abusive conditions. The sintered electrode cells do have a characteristic termed “memory effect.” This arises if the cell has an exact repetitive charge-discharge regime. After a period of time, if the cell is asked to discharge longer than the specified regime, it loses voltage as if it “remembers” the discharge regime. The cell is not damaged. Discharge can continue at a lower voltage and the “effect” eliminated by reconditioning with a series of full capacity charge-discharges.

Nickel Cadmium Pocket and Tubular Plate Cells

Materials fabrication of the Ni-Cd pocket plate and tubular cells follow essentially an identical process to that described for the Ni-Fe cell construction but with cadmium substituted for the iron in the negative electrode construction. The preparation of the cadmium active material uses a co-precipitation process, usually starting with a CdSO_4 or CdCl_2 solution where KOH is added to precipitate the $\text{Cd}(\text{OH})_2$ used in cell fabrication. Here, graphite is used to provide internal

conductivity to contact the positive nickel active mass. On overcharge, the oxygen evolution attacks the graphite conductor to produce CO_2 which, in turn, reacts with the KOH electrolyte to produce K_2CO_3 , lowering the concentration of the electrolyte. Each overcharge causes loss of graphite and decreases the available contact to the nickel active mass.

Because they can withstand overcharge, overdischarge, mechanical and electrical abuse, and require little maintenance, the Ni-Cd batteries find use in critical industrial applications including switchgear operations, telecommunications, uninterruptible power supplies (UPS), and emergency lighting even though their cost is higher than the lead acid battery system. The Ni-Cd vented pocket plate battery was used in the world's largest UPS power supply in Anchorage, Alaska.

Nickel Cadmium Fiber Plate Batteries

Fiber Plate Electrode Technology was developed by DAUG in Germany in 1983. The electrodes have an open three-dimensional nickel fiber structure that provides good conductivity throughout the electrode structure to ensure excellent electrical performance. The fibrous nickel matrix current collector matrix forms a lightweight flexible structure with about 90% of the electrode volume available for holding the active material. It is free from graphite and iron, found in other nickel-based cathode battery systems; so reduces water consumption by electrolysis and eliminates carbonation of the electrolyte on overcharge to ensure very long cycle life.

The elastic nature of a NCFP electrode and its fibrous current collector construction enable it to accommodate the volume changes and maintain good contact with the active materials during the charge and discharge cycles, as well as absorb shock and vibration. This feature results in longer cycle life compared to other Ni-Cd battery technologies. The NCFP technology is considerably lighter in weight and has higher energy density than the regular Ni-Cd constructions. Fiber plate batteries can be designed with very thin plates for ultra high rate discharge performance or with thick plates for long duration lower current discharge applications. Other features include no need for change of electrolyte throughout life due to graphite-free electrodes, low water consumption, long shelf life, and flame-arresting vent protection.

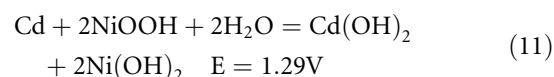
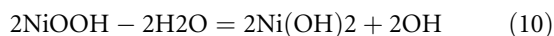
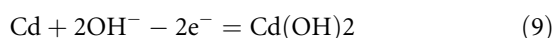
Applications include switchgear protection, emergency lighting, motive power, train lighting, instrumentation and process control, UPS, electric vehicles, diesel engine cranking, aircraft and helicopter ground starting.

Nickel Cadmium Sealed Sinter Plate Cells

The present sealed Ni-Cd sinter plate battery technology evolved from the sealed battery technology developed by Varta in the 1940's based on an IG Farben development in the 1930's. The sinter plate construction, developed in the 1940 time frame, has low internal resistance, superior high rate and low temperature performance compared to the pocket plate cell design. The batteries are sealed and can be used in any orientation. This battery powered the Luftwaffe in the 1940s, the early flights in the space program and in the development of portable electronic devices, such as cellular phones and portable notebook computers.

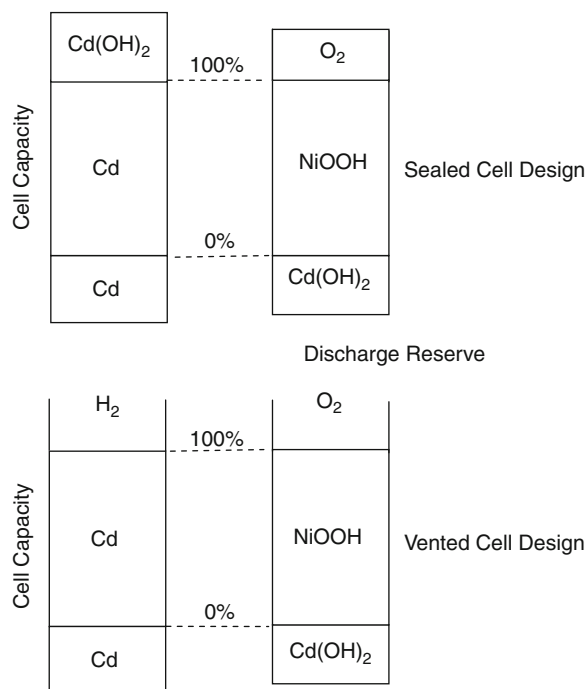
The small sealed sintered plate Ni-Cd cylindrical and prismatic cell constructions have the favorable characteristics of the pocket/tubular plate constructions with increased energy density. The electrodes are thin and the cells have low internal resistance with excellent high rate and low temperature performance. The stem is rugged and can withstand physical and electrical abuse. It requires minimal maintenance.

Advantages	Disadvantages
Sealed, no maintenance	Higher cost than lead acid
Long cycle and shelf life	Lower capacity than Li-Ion and Ni-MH
Excellent high rate performance	Environmental issues with cadmium
Flat discharge profile	
−40°C to +50°C operating range	
Operate in any orientation	



The electrodes for cell fabrication are produced by vacuum impregnation of the nickel sinter plate. The sinter plates are produced by coating slurry of carbonyl nickel powder in CMC onto a nickel-plated steel or nickel screen current collector and burn off the CMC at high temperature. This process produces a nickel current collector with about 80% porosity. The coils or plates (depending on cell construction) of the sinter plate are vacuum impregnated in an aqueous nickel nitrate solution, dipped into a NaOH bath to convert the nickel nitrate into $\text{Ni}(\text{OH})_2$, and then dried. The process is repeated until the desired loading of $\text{Ni}(\text{OH})_2$ is obtained. The negative electrode is produced by a similar process except cadmium nitrate solution is used to impregnate the sinter electrode. The impregnated strips of cobalt and nickel electrode stock are cut to size, stacked as plates or wound into a cylinder, and assembled in the appropriate cell container.

The sealed cylindrical and prismatic cells can operate in any orientation. They were instrumental in the electronic revolution that produced the portable computers, cellular phones, digital cameras, and the like. This requires specific design features that prevent pressure buildup from oxygen gas generation on overcharge, with no free electrolyte and no maintenance or servicing. The internal design of each electrode of the cell is critical to long life at full performance. The cell balance, ratio of the capacity of Cd to the capacity of NiOOH electrodes, is carefully controlled to prevent any hydrogen evolution from the cadmium electrode. Sealed cells have a different balance (ratio of the negative and positive electrode capacities) from that for vented cell constructions. In sealed Ni-Cd cells, it is essential to prevent hydrogen evolution on charge as it causes the cell to dry out and develop high internal pressure. This is accomplished by balancing the capacity of each electrode, positive and negative, during cell assembly to ensure that oxygen is the only gas evolved, as depicted in Fig. 1. As the cells reach full charge and enter overcharge, oxygen is evolved from the nickel positive. The oxygen diffuses through/around the porous separator to the cadmium electrode where it reacts directly with the active cadmium metal, just deposited on the negative electrode, to produce $\text{Cd}(\text{OH})_2$ with heat evolution. To that end, the negative cadmium plate is designed with an excess of uncharged active mass.



Nickel-Based Battery Systems. Figure 1

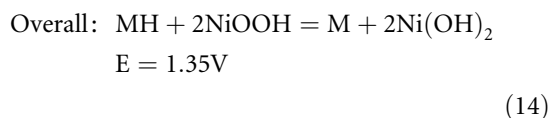
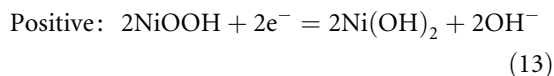
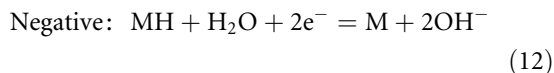
Cell balance for vented and sealed nickel cadmium cells. It is important to avoid hydrogen gas evolution in the sealed cell

Nickel Metal Hydride

Nickel metal hydride (Ni-MH) batteries have application in hybrid electric vehicles, portable electronic devices such as cameras, shavers, toothbrushes, etc. They offer a higher energy option than Ni-Cd batteries with the hydrogen-absorbing metal hydride alloy (MH) replacing the cadmium electrode in Ni-Cd cell construction. Like Ni-Cd, the Ni-MH also shows the “memory effect” on repetitive cycle regimes. The nickel metal hydride battery was introduced commercially in 1989. The technology is based on the development of rare earth alloys with nickel that have the ability to reversibly absorb and desorb hydrogen. The nickel metal hydride (MH) electrode replaces the cadmium electrode in the Ni-Cd cell construction. It stores, and can deliver, more energy than an equivalent-sized Ni-Cd cell.

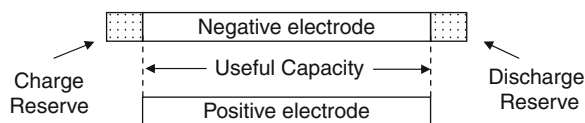
The Ni-MH battery has found wide application as a replacement for rechargeable Ni-Cd cells and primary alkaline manganese cells, especially in portable

electronic devices. It offers a rechargeable replacement for the primary alkaline cell and gives superior performance. Although it initially has a lower voltage, it has a constant voltage discharge as compared to the sloping voltage of the alkaline cell that is very desirable for cameras, portable electronics, etc. More recently, it has found application in hybrid electric automobiles.



The “metal” M in the equations above refers to the negative electrode of a Ni-MH cell and is an intermetallic alloy or compound capable of storing and releasing hydrogen. These compounds can store many times their volume of hydrogen. The hydrogen atom is much smaller than nickel and can occupy interstices in the nickel alloy anode. In theory, the MH alloy has almost twice the capacity of the cadmium anode material. This gives the Ni-MH battery the capability of about 20% greater energy storage than the Ni-Cd battery. Many different compounds have been developed for this application. The most common is AB₅, where A is a mixture of rare earth elements (chromium, lanthanum, cerium, neodymium, and praseodymium) and B is a mixture of nickel, cobalt, manganese, or aluminum. The metal hydride electrode substitutes for the cadmium electrode in the Ni-Cd cell and makes for a more environmentally friendly system. Many design parameters are similar to those in Ni-Cd.

Alloys used for the negative electrode of Ni-MH secondary batteries function as a hydrogen storage media as well as a catalyst for the hydrogen electrode reaction. The hydrogen-absorbing alloys can be divided into three main types: the AB₅, A₂B₇, and AB₂ types. In the LaNi₅ alloys, La is substituted for misch metal, Mm, a mixture of rare earth elements including La, Ce, Pr, and Nd and partially substituted for Ni for Co and Al to increase the hydrogen storage and lengthen the charge-discharge cycle life. The cell balance of the capacity of the positive and negative electrodes for best performance are given in Fig. 2.



Nickel-Based Battery Systems. Figure 2
Cell balance in sealed nickel metal hydride cells

Ni-MH cells can be practically charged at 1 C rate using a constant current voltage charge regime and discharged at 4 C. Its 1 C charge-discharge cycle life expectancy is more than 500 cycles enabling commercial applications of Ni-MH to a wide variety of cordless equipment. The Ni-MH cells do not have the memory effect characteristic, where cells discharged on the exact same charge-discharge regime “memorize” the cycle and if discharged, further, act as if fully discharged.

As with Ni-Cd, when overcharged at low rates, the oxygen produced at the positive electrode on overcharge diffuses through the porous separator and recombines at the surface of the negative to suppress hydrogen evolution and converts the charging energy into heat. This process prevents internal pressure buildup to allow Ni-MH cells to remain sealed if overcharged during normal operation. Ni-MH cells use a KOH electrolyte and the same hydrophilic polyolefin nonwoven separator materials used in the Ni-Cd cell.

The internal construction of the Ni-MH cell mimics that of the Ni-Cd cells. The Ni-MH cells are available in a wide variety of sizes ranging from the AA-size for digital cameras, up to 50 Ah or larger used in hybrid car applications. Useful discharge capacity is constant up to the 1 h discharge rate and does not differ significantly from the rated capacity.

Nickel metal hydride batteries consist of a positive electrode containing a mixture of carbon/graphite conductive diluent and nickel hydroxide as its principal active material. The negative electrode consists mainly of hydrogen-absorbing conducting metal alloys, a porous polymer separator filled with KOH electrolyte, a metal case and a resealing vent. The anode-separator-cathode foils are wound in a cylinder or stacked in a prismatic format and inserted into a nickel-plated steel can. The prismatic cell construction may also be formed by stacking alternate positive and negative individual electrodes or by winding the electrode strips into a cylindrical form and then

inserted into a prismatic can or wound and inserted into a cylindrical can construction.

A complete discharge of a Ni-MH cell causes permanent damage to the negative electrode. This situation can occur in a group of series-connected cells if one low-capacity cell in the series is completely discharged and driven into reverse due to its lower capacity than surrounding cells. When this happens, the good cells will start to drive the discharged cell in reverse and can cause permanent damage to that cell.

Ni-MH cells have a 5–10% self-discharge rate at the beginning of life but stabilize around 0.5–1% per day. Cooler storage temperatures have lower discharge rate and longer battery life. Continual trickle charging of small Ni-MH cells can cause permanent deterioration in cell performance.

Ni-MH batteries provide a relatively constant voltage during the discharge cycle, unlike a standard alkaline battery where the voltage falls steadily during discharge. This characteristic gives the Ni-MH longer useful service to power portable electronic devices than the standard alkaline zinc manganese cell and can be substituted for the AA-size alkaline cell in most devices.

Ni-MH batteries have application in hybrid electric vehicles and portable electronic devices, such as cameras, shavers, toothbrushes, etc. The Ni-MH battery is also used in hybrid cars to improve stop-start and acceleration performance. It absorbs braking energy to assist the braking operations and during the acceleration, returns that energy to supplement the gas engine.

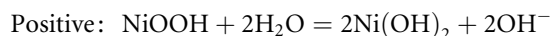
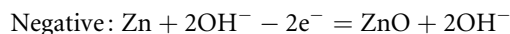
Improper disposal of Ni-MH batteries poses less environmental hazard than that of Ni-Cd because of the absence of toxic cadmium.

Nickel-Zinc Batteries

The performance of nickel-zinc (Ni-Zn) batteries lies between that of high-energy silver-zinc and nickel cadmium. The nickel-zinc cells have excellent high rate performance and about twice the energy density of Ni-Cd but the cycle life is severely limited due to “shape change” problems that limit recharge of the zinc electrode. While Junger and Edison were both awarded patents on the rechargeable Ni-Zn battery, the first commercial cell was developed by J. Drumm in Dublin, Ireland, in 1931 to power the lights on a passenger train between Bray and Dublin. It ran

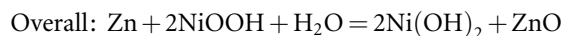
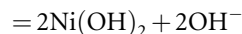
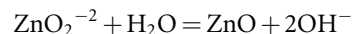
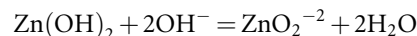
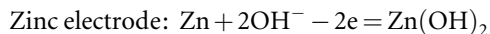
until 1949. Later, Yuasa introduced a 10 Ah sealed Ni-Zn battery for use in lawn mowers and light duty electric vehicles without success. Their cells limited cycle life of 200 cycles at the two-hour rate and were not sufficient for a commercial rechargeable battery system. PowerGenix successfully demonstrated that the Ni-Zn battery could substitute for the Ni-MH battery in a Toyota Prius car. A potentially new large application for Ni-Zn is stationary energy storage.

Although the zinc electrode is a difficult system and can deliver excellent high rate performance, it suffers from problems with the surface morphology, dendrite formation on charge, shape change, agglomeration/redistribution and passivation, etc., that limit cycle life. The problems directly result from the characteristics of the reactions of the zinc ions in solution and the reactivity of zinc metal with the electrolyte.



$$E = 1.73\text{V}$$

In 2008, PowerGenix introduced a new sealed cell for the power tool market with a modified KOH electrolyte formulation. The higher voltage of the zinc electrode results in a unit cell voltage of 1.73 versus 1.2 for Ni-Cd, giving the Ni-Zn a distinct advantage in energy density. In addition, the facile kinetics of the zinc electrode gives the system superior performance for high rate applications. The reaction of the active materials with the electrolyte limits the cell life.



Cell construction for the Ni-Zn cell is similar to that for the sealed spiral wrap cylindrical and prismatic Ni-Cd and Ni-MH cells for portable electronic applications but with some differences. The Ni-Zn cell incorporates a multilayer micro-porous separator

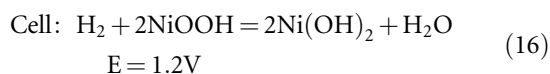
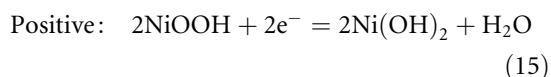
with “wicking” components used to maintain uniform electrolyte distribution, as well as acting as a buffer/barrier to prevent shorting from zinc dendrite formation. The barrier also slows oxygen gas diffusion and lowers the recombination rate that can lead to pressure buildup inside the cell. The KOH electrolyte contains CaO additions to calcium zincate that can control the free zinc ions in solution. In the cell electrolyte, the CaO reacts with zinc ions to form calcium zincate. Other improvements center on the use of fiber (steel wool) current collectors for the zinc electrode. The Ni-Zn cells have excellent high rate and low temperature performance.

Nickel-Hydrogen Batteries

The nickel-hydrogen (Ni-H₂) battery was developed in the late 1960s as a new, long-lived, highly reliable power source to replace Ni-Cd for space applications. It won out over strong competition from regenerative fuel cells mainly due to issues related to life expectancy and system engineering problems in that resulted in increased size, weight, complexity, loss of catalytic activity, and cost to assure the required long life.

The Ni-H₂ system characteristics of its lightweight, extremely long cycle life and simple, yet, accurate state of charge determination based on a pressure measurement, makes it an ideal system for all space applications ranging from low orbits, mid-altitude orbits, to the higher geosynchronous orbits. It replaced the Ni-Cd system because of its lighter weight and longer life. It has powered the space shuttle, space stations, and interplanetary probes including extending planetary exploration stations on the moon and Mars. Ni-H₂ battery cells are by far the longest-lived and most reliable of all sealed rechargeable battery systems.

Advantages	Disadvantages
Long calendar life	Expensive
10+ year's cycle life	Low energy density
Accurate state-of-charge measure	High self-discharge
Abuse-tolerant	



Internal cell design considerations include the depth of overdischarge, positive pre-charge, or negative pre-charge cell designs. The type of pre-charge depends on if there is excess capacity associated with the positive nickel electrodes (positive pre-charge) or the negative hydrogen material (negative pre-charge). In this configuration, there is a capacity loss mechanism caused by the reaction of hydrogen with the cobalt additive in the active material of the positive electrodes. Newer cell designs specify positive pre-charge designs where there is capacity remaining in the positive electrodes of the cell at the end of discharge.

The hydrogen electrodes are Teflon bonded, platinum black catalyzed, supported on a nickel substrate. A high surface area platinum black hydrogen catalyst loading, in the range of 7 mg/cm², ensures long life. The electrodes have a porous Teflon backing to stop water loss from the back side while allowing full access for gas transport. The system is designed so that on overcharge, O₂ is forced out of the backside of the electrode to prevent direct contact with the active surface of the hydrogen electrode. Usually, a fibrous Zircar nonwoven ceramic separator holds the KOH electrolyte.

The sintered nickel substrate for the cathode is similar to that used for Ni-Cd and Ni-MH cell constructions. The nickel active materials are loaded into the sinter plate using either an aqueous (Bellcore) or alcoholic-based (Air Force or Pickett) electrochemical impregnation process. A 5–10% cobalt additive is deposited with nickel hydroxide to improve charge acceptance. These electrodes have a significantly longer cycle life over the standard vacuum impregnated or pasted nickel electrodes used in commercial Ni-Cd cells.

In the “pineapple” cell design, two of these nickel sinter electrodes are stacked back-to-back with asbestos cloth separators on each face of the electrodes. The cell construction is named after an exploded view of the slices (electrode components) in a can of pineapple. The recirculating stack design permits a natural concentration gradient across the stack. A thin flame-sprayed zirconium oxide coating on the inner wall of the pressure vessel is contacted by the protruding edges of the separator to redistribute the electrolyte through the diffusive gradient set up by the recirculation

pattern. The hydrogen electrodes are platinum-catalyst-based electrodes with a hydrophilic platinum catalyst on the side facing the separator, and the side facing the gas screen consisting of a hydrophobic porous Teflon layer. The electrolyte is 31% by weight aqueous potassium hydroxide. The cell components are stacked on a polysulfone core that fits into the center hole of the pineapple slice components.

All nickel-hydrogen cell designs have the electrolyte contained within the porous stack components. A hydrophilic wall wick provides a path for diffusion of the electrolyte by capillary forces from wet to dry regions of the stack. The finished Ni-H₂ cell is a sealed pressure vessel with positive and negative terminals.

The second-generation cell, 5.5 in. diameter cell design, has capacities up to 350 Ah. Eagle Picher developed a cell line based on the electrochemical impregnation method for nickel electrodes employing either slurry sinter or dry sinter. The nominal thickness of the nickel electrode was 35 mils, as opposed to the 30-mil thickness of the electrodes used in the COMSAT cells. The cells were built with a back-to-back stacking arrangement that included two layers of Zircar separator, using a 31% potassium hydroxide electrolyte. They were built with about a 15% nickel pre-charge. A nylon compression seal was used to seal the terminals into the pressure vessel. A flame-sprayed zirconium oxide wall wick was included in the design to allow electrolyte gradients to equilibrate by diffusion. The cells had no free electrolyte.

A dual-stack cell design that enclosed two stacks within the same pressure vessel was developed to reduce the number of pressure vessels required in some power systems. This configuration was termed a common pressure vessel. These powered the Iridium constellation of 88 satellites (98 satellites, including orbital spares). While these batteries have functioned reasonably well, there are some difficulties in maintaining all the cells at a similar state of charge. However, these single pressure vessel batteries continue to perform well in orbit.

Bibliography

1. Anderman M, Baker C, Cohen F (1997) Proceedings of the 32nd intersociety energy conversion conference, Honolulu, Hawaii, vol 1, p 97465 Honolulu, Hawaii
2. Anderson B, Ojefors L (1979) In: Thompson JF (ed) Power sources, vol 7. Academic, London, p 329
3. Baker C, Barekatien M (2000) Proceedings of the SAE power systems conference, San Diego, 2000
4. Beauchamp RL (1971) US Patent 3,573,101
5. Beauchamp RL (1972) US Patent 2,653,967
6. Beccu K (1972) US Patent 3,669,745
7. Buzzelli E (1978) Silver-iron battery performance characteristics. In: Proceedings of the 28th power sources symposium, Electrochemical Society, Pennington, p 160
8. (2004) Cadnica sealed type Nickel-Cadmium batteries engineering handbook, Sanyo Electric Company, Osaka
9. Casellato U, Comisso N, Mengoli G (2006) Effects of Li ions on reduction of Fe oxides in aqueous alkaline medium. *Electrochimica Acta* 51:5669–5681
10. Cook J (1999) Separator-hidden talent Electric and hybrid vehicle technology
11. Corrigan DA, Venkatesan S, Gifford P, Holland A, Fetcenko MA, Dhar SK, Ovshinsky SR (1997) Proceedings of the 14th international electric vehicle symposium, Orlando
12. Corrigan DA, Knight SK (1996) *J Electrochem Soc* 143(5):1613
13. Ettel V, Ambrose J, Cushnie K, Bell JAE, Paserin V, Kalal PJ (1997) US Patent 5,700,363
14. Falk SU, Salkind AJ (1969) Alkaline storage batteries. Wiley, New York
15. Feduska W, Rosy R (1980) An advanced technology Iron-Nickel battery for electric vehicle propulsion. In: Proceedings of the 15th IECEC, Seattle, p 1192
16. Fetcenko MA, Ovshinsky SR, Chao B, Reichman B (1996) US Patent 5,536,591
17. Fetchenko MA, Venkatesan S, Ovshinsky S (1991) Proceedings of the symposium on hydrogen storage materials, batteries and electrochemistry, Electrochemical Society, Pennington, p 141
18. Fleischer A (1948) *J Electrochem Soc* 94:289
19. Ford FE (1994) Handbook for handling and storage of Nickel-Cadmium batteries: lessons learned, NASA Ref. Publ. 1326
20. Gutjahr MA, Buchner H, Beccu KD, Saufferer H (1973) In: Collins DH (ed) Power sources, vol 4. Oriel, Newcastle Upon Tyne, p 79
21. Halpert G (1984) *J Power Sources* 12:117
22. Halpert G (1990) Proceedings of the symposium on nickel hydroxide electrodes, The Electrochemical Society, Hollywood, Oct 1989, *J Electrochem Soc*, Pennington, pp 3–17
23. Hill TE, Rosy R, Vaill RE (1978) Performance characteristics of iron nickel batteries. In: Proceedings of the 28th power sources symposium, Electrochemical Society, Pennington, p 149
24. Hudson R, Broglio E (1980) Development of nickel-iron battery system for electric vehicle propulsion. In: Proceedings of the 29th power sources conference, Electrochemical Society, Pennington
25. Ishiwa K, Ito T, Miyamoto K, Takano K, Suzuki S (1999) Evolution and extension of NiMH technology. In: 16th international seminar on primary and secondary batteries, Ft. Lauderdale
26. Kanagawa I (1998) 15th international seminar on primary and secondary batteries, Ft. Lauderdale
27. Kruger FJ (1998) 15th international seminar on primary and secondary batteries, Ft. Lauderdale

28. Kulin TM (1998) 33rd intersociety engineering conference on energy conversion, ICECE-98-145, Colorado Springs, 2–6 Aug 1998
29. Reddy TB (2011) Linden's handbook of batteries, 4th edn. McGraw Hill, New York
30. Lindstrom O (1975) In: Collins DH (ed) Power sources, vol 5. Academic, London, p 283
31. McBreen J (1990) The nickel oxide electrode. In: White RE, Bockris JO'M, Conway BE (eds) Modern aspect of electrochemistry, vol 21. Plenum, New York, p 29
32. McRae B, Nary D (1998) Proceedings of the 38th power sources conference, pp 123–126
33. Matsumoto I, Ogawa H, Iwaki T, Ikeyama M (1988) 16th international power sources symposium, 1988, Bournemouth, England
34. Mil-B-81757 (1984) Performance specifications, batteries and cells, storage, nickel cadmium, Aircraft General Specification, Crane Division, NSWC, 1 July 1984
35. Mishima R, Miyamura H, Sakai T, Kuriyama N, Ishikawa H, Uehara I (1993) J Alloys Compd 192:176–178
36. Notten PHL, Hokkeling P (1991) J Electrochem Soc 138(7):1877
37. Notten PHL, Daams JLC, Einerhand REF (1992) Ber Bunsenges Phys Chem 96:5
38. Nickel-Cadmium batteries (2004) charge system guide, Panasonic Industrial Company, Secaucus
39. Ohta K, Matsuda H, Ikoma M, Morishita N, Toyoguchi Y (1996) US Patent 5,571,636
40. Ojefors L, Carlson L (1977) An iron-air vehicle battery. J Power Sources 2:287
41. Oshitani M, Yufu H, Takashima K, Tsuji S, Matsumaru Y (1989) J Electrochem Soc 136:6
42. Oshitani M, Yufu H (1989) US Patent 4,844,999
43. Ovshinsky SR, Dhar SK, Fetcenko MA, Young K, Reichman B, Fierro C, Koch J, Martin F, Mays W, Sommers B, Ouchi T, Zallen A, Young R (2000) 17th international seminar and exhibit on primary and secondary batteries, Ft. Lauderdale, 6–9 Mar 2000
44. Ovshinsky SR (1998) Materials research society fall meeting, Boston
45. Ovshinsky SR, Fetcenko M, Ross J (1993) Science 260:176
46. Ovshinsky SR (1991) In: Adler D, Schwartz B, Silver M (eds) Disordered materials: science and technology. Institute for Amorphous Studies Series/Plenum Publishing Corporation, New York
47. Ovshinsky SR, Corrigan DA, Venkatesan S, Young R, Fierro C, Fetcenko M (1994) US Patent 5,348,822, 14 Apr 1994
48. Pell MB, Blossom RW (1970) US Patent 3,507,699
49. Picket DF, Maloy JT (1978) J Electrochem Soc 12:1026
50. Picket DF (1974) US Patent 3,827,911
51. Picket DF (1975) US Patent 3,873,368
52. Puglisi V (2000) 17th international seminar and exhibit on primary and secondary batteries, Ft. Lauderdale 6–9 Mar 2000
53. Reichman B, Mays W, Fetcenko MA, Ovshinsky SR (1999) Electrochemical society proceedings, vol 97–16, Oct 1999
54. Salkind AJ, Venuto CJ, Falk SU (1964) The reaction at the iron alkaline electrode. J Electrochem Soc 111:493
55. Sapru SR, Reichman B, Reger A, Ovshinsky SR (1986) US Patent 4,623,597
56. Singh D, Wu T, Wendling M, Bendale P, Ware J, Ritter D, Zhang L (1998) Mater Res Soc Proc 496:25–36
57. Souza A, Carlos IA, Lopes M, Finazzi GA, de Almeida MRH (2004) Self-discharge of Fe-Ni alkaline batteries. J Power Sources 132:288–290
58. Stempel RC, Ovshinsky WR, Gifford PR, Corrigan DA (1998) IEEE spectrum 35(11):29–34
59. Takagi S, Minohara T (2000) Society of automotive engineers, 2000-01-1060, Mar 2000
60. Tuomi D (1976) The forming process in nickel positive electrodes. J Electrochem Soc 123:1691
61. van Beek JR, Donkersloot HC, Willems JJG (1984) Proceedings of the 14th international power sources symposium (1984), Waikola, Hawaii
62. Watanabe K, Koseki M, Kumagai N (1996) J Power Sources 58:23–28
63. Weininger JL (1982) In: Gunther RG, Gross S (eds) The nickel electrode, vol 82–84. Electrochemical Society, Pennington, pp 1–19
64. Weizhong T, Guangfei S (1994) J Alloy Compd 203:195–198
65. Young K, Fetcenko MA, Reichman B, Mays W, Ovshinsky SR (2000) Proceedings of the 197th electrochemical society meeting, May 2000
66. Yu X, Licht S (2007) Advances in Fe(VI) charge storage part I, primary alkaline super-iron batteries. J Power Sources 171:966–980
67. Yu X, Licht S (2007) Advances in Fe(VI) charge storage part ii, reversible alkaline super-iron batteries and nonaqueous super-iron batteries. J Power Sources 171:1010–1022

Night Vision Pedestrian Warning in Intelligent Vehicles

MATTHIAS SERFLING, OTTO LÖHLEIN

Machine Perception Systems and Reliability,
Environment Perception Daimler AG, Ulm, Germany

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Discussion of Fusion Concepts
System Overview
Sensor Configuration
Boosted Feature-Level Sensor Fusion
Global Positioning and Road Course Estimation
Threat Analysis

Experiments
 Conclusion
 Future Directions
 Acknowledgments
 Bibliography

Glossary

ACC Adaptive Cruise Control – A cruise control system which controls the vehicle speed in order to keep a set distance to the vehicle in front.

AdaBoost Adaptive Boosting – A supervised machine learning algorithm which combines several weak learners (very simple classifiers) to construct a strong learner classifier.

Baseline The baseline of a stereo camera system defines the distance of the corresponding camera coordinate systems and influences the accuracy of the depth analysis.

Bayes classifier A simple probabilistic classifier based on applying Bayes' theorem from Bayesian statistics.

CAN Controller Area Network – A standard vehicle bus system which is designed to allow microcontrollers and devices to communicate with each other within a vehicle.

CFAR Constant False Alarm Rate – An adaptive algorithm used especially in radar systems to detect targets in noisy measurements.

CMOS Complementary Metal Oxide Semiconductor – A technology for constructing integrated circuits.

CTRV Constant Turn Rate and Velocity – A motion model which assumes an object moves in sample time T with constant velocity and constant turn rate.

Disparity A term in multiple view geometry in computer vision which describes the distance of two points in two different images by identifying the same point in the scene observed from different positions.

Digital map The knowledge database used in automotive navigation systems which provide information about the road network.

Doppler An effect that describes the change in frequency of a wave for an observer moving relative to the source of the wave.

EU European Union

Epipolar line A line describing the positions in a sensor coordinate system, which correspond to a single position in a second sensor coordinate system. Typical term used in stereo camera systems.

False alarm rate A failure measure in detection systems describing the amount or the rate of detections which do not correspond to a proper object.

FIR (far-infrared) A device analyzing the infrared electromagnetic radiation with a wavelength between 0.7 and 300 μm .

Flat world assumption A simple model in the field of computer vision which assumes that all objects perceived by a sensor (e.g., a camera) are positioned in a flat world.

GPS (global positioning system) A space-based global navigation satellite system that provides reliable location and time information of an object on or near the Earth.

Haar wavelets A basic wavelet method to decompose a signal into independent signal parts.

INS Inertial Navigation System – A system that uses motion sensors and rotational sensors installed on a platform to continuously calculate via dead reckoning the position, orientation, and velocity of a moving object without any external references.

MCL Monte Carlo Localization – A method in robotics and sensors to determine the position of a robot given a map of its environment based on Markov localization.

NIR Near-Infrared – A device analyzing the electromagnetic spectrum from approximately 800 nm to 2,500 nm.

NIRWARN Near-Infrared Warning

Particle Filter A probabilistic method in the field of computer vision to track objects using a Monte Carlo approach.

PREVENT Preventive and Active Safety Application, European founded project.

ProFusion ProFusion Subproject of the PREVENT project.

RADAR Radio Detection and Ranging – An object-detection system that uses an electromagnetic wave to identify the range, direction, (and velocity) of a target.

ROI Region of interest – A small region selected for further processing.

ROC Receiver operating characteristic – A graphical plot to show the fraction of true positives vs. the fraction of false positives for a binary classifier system as its discrimination threshold is varied.

Stereo camera system A type of camera system which observes the environment with two lenses in order to simulate human binocular vision, and thus reconstructs the depth information of the environment.

Two-class problem A typical classification problem with an object class and a non-object class.

UTM Universal Transverse Mercator – A grid-based two-dimensional Cartesian coordinate system, which specifies locations on the surface of the Earth.

VGA Video Graphics Array – A standard VGA camera has a resolution of 640×480 pixels.

Definition of the Subject and Its Importance

Today's second generation of automotive night vision systems uses near-infrared (NIR) or far-infrared (FIR) cameras combined with integrated pedestrian detection systems. These systems direct the driver's attention by showing all detected pedestrians on a display inside the car. This enables the driver to analyze an imminent situation earlier and react appropriately to prevent a potential hazardous situation, but the driver has to decide on his own whether a detected pedestrian is a potential risk or not. Thus, the next generation of night vision systems needs to integrate an additional warning component which signals to the driver only the relevant objects, that is, all objects on the road or near the road. Such a warning system needs to detect objects more reliably and sooner and has to provide information about the object's position. Moreover, the system has to know the road course in front of the vehicle to do a threat analysis of the detected objects by their position. According to [1] and [2], the brake reaction time under real-life conditions is about 1.5–2 s. Incorporating also the braking time to standstill at a common speed of 100 kph on rural roads, the driver needs to be warned at least 6 s in advance.

In this context, a reliable and robust pedestrian detection system as well as a road course estimation system which covers both factors at a distance up to 120 m is essential for the development of the next generation of night vision warning system.

As shown in a German funded project NIRWARN [3], single NIR sensor systems can provide pedestrian detection reliably up to 90 m. For the enlargement of the detection range to 120 m, the illumination of the road and the camera resolution and sensitivity are limiting factors. Neglecting the cost factor of an automotive FIR camera, which can achieve these ranges, the problem of providing information about the object's position still remains with any kind of single camera sensor system. In fact, a stereo camera system can provide position information, but a large baseline (i.e., at least 0.8 m using a VGA imaging sensor with lenses of focal length 12 mm) is required to realize detection ranges of 120 m. A large baseline is difficult to integrate into a passenger car.

The known approaches for lane or road recognition systems perform well under daylight conditions on motorways [4–9], but these systems mostly fail or have a very limited prediction horizon on rural roads during the night due to high curvature road types, missing lane markings, heavily changing lighting conditions, and varying road surfaces.

This contribution addresses both problems of a reliable and robust pedestrian detection system with enlarged detection range, and also provides information about the object's position and a reliable road course estimation system covering distances up to 120 m in order to carry out a threat analysis and to signal to the driver only the relevant objects. To achieve the full functionality, the presented approach combines the sensor information provided by a NIR camera, an imaging radar, and a digital map.

Unlike other mid- and high-level radar camera fusion approaches [10–15], the proposed system realizes an early feature-level fusion to detect pedestrians at night within 120 m in rural and suburban environment. The presented low-level fusion system shows better results than a camera-only system, outperforms a fusion system on higher levels, and manages the higher computational load of processing raw data.

In order to overcome the drawback of the visibility constraints of an image sensor, a digital map in combination with a new highly sensitive imaging radar is used to provide a reliable prediction of the road course. Unlike an image sensor, the used imaging radar sensor is capable of clearly detecting the road boundary independent of lane markings and for the most part

independent of weather and lighting conditions at least up to 60 m. Reflections from trees, posts, traffic signs, parked cars, and guardrails enlarge the detection range of the road boundary up to 100 m.

Introduction

Many approaches for pedestrian recognition under daylight conditions for vision systems have been developed, especially for shorter detection ranges. Single-sensor pedestrian recognition at night based on image processing is widely discussed in [16–19]. The preferred detection method applied to near-infrared (NIR) images [16, 17] and far-infrared (FIR) images [18, 19] is the cascaded classification using Haar wavelets. Note that [19] uses a pixel classification prestage to reduce the input space and uses in addition to the cascaded classifier, a second classifier based on object shape matching. A particle filter fuses the detection results from both classifiers.

On the other hand, multi-sensor pedestrian recognition has been presented in [20] and [21]. A two-step fusion of a radar sensor and monocular vision sensor is applied in [20]. The first step generates a list of potential targets using the radar sensor. The second step utilizes the images from a vision sensor to verify the targets, applying flexible shape models trained from manually extracted pedestrians. Thus the false alarm rate can be reduced analyzing the image in small regions of interest (ROI), but the detection rate is strictly limited by the radar sensor.

A three-level early fusion system for road user detection including cars and pedestrians is presented in [21]. The fusion process combines raw sensor data provided by a laser scanner, a long and a short range radar, a gray-scale camera, and a FIR camera at an early stage of the data processing chain and performs a joint data interpretation with respect to a common model basis. The first step generates hypotheses for each sensor and merges overlapping hypotheses of different sensors in one hypothesis and assigns the involved sensor measurements. In the second fusion level, different sensor features (e.g., velocity, object dimensions, and FIR image edges) are used to classify the hypotheses. Finally, the third fusion level takes into account the assigned measurements (e.g., object position) of each hypothesis to track the object.

This contribution presents a robust and reliable system for the recognition of pedestrian at night. In contrast to the fusion techniques above, the developed detection system fuses raw sensor data using a boosted classifier. In order to reduce additional costs for new sensors, fusion approach is applied to a scanning ACC radar sensor and an NIR camera sensor, commonly available in modern cars.

The next section gives a general overview of fusion concepts. Moreover, the advantages of feature-level fusion compared with a high-level fusion approach are discussed based on a simple experiment. The developed application of a pedestrian warning system is presented in sections “[System Overview](#)” to “[Threat Analysis](#).” First, a system overview gives a short description of all components. Section “[Sensor Configuration](#)” describes the sensor configuration and the calibration algorithm to associate measurements from different sensors. Section “[Boosted Feature-Level Sensor Fusion](#)” presents the developed feature-level fusion approach to detect pedestrians reliably and robustly with an enlarged detection range. The developed road course estimation subsystem is described in section “[Global Positioning and Road Course Estimation](#).” The last subsystem, discussed in section “[Threat Analysis](#),” combines the two subsystems to realize a threat analysis of detected pedestrians depending on their position on the road. Section “[Experiments](#)” analyzes the performance and the results of these subsystems. Finally, section “[Conclusion](#)” provides a conclusion and section “[Future Directions](#)” discusses the impact on the development of upcoming fusion systems.

Discussion of Fusion Concepts

The special requirements for an automotive pedestrian recognition system can be achieved by means of two strategies. First, enhance the sensing performance of the sensors used in terms of higher spatial resolution and sensitivity and an improved recognition technology. Second, exploit synergistically the electromagnetic spectrum by fusion of different physical sensor technologies.

Sensor fusion allows exploiting the best of all sensors to overcome the shortcomings of one individual sensor. This provides for an enriched knowledge about

the environment in comparison to only one sensor or sensor technology.

The EU projects ProFusion 1 and ProFusion 2 [22] investigated five multiple-sensor data fusion approaches for safety applications, namely, early fusion (also named feature-level fusion), track-level fusion, multilevel fusion, feedback fusion, and grid-based fusion.

Early fusion techniques combine data provided by multiple sensors at an early stage of the data processing chain and perform a joint data interpretation on aggregated data with respect to a common model basis. Two different models for early fusion are possible: pixel-level fusion, only applicable for fusion of images with the same mapping rule, and feature-level fusion, also adaptable for arbitrary sensor combinations. Early fusion techniques demand extensive data processing, which directly requires more complex hardware, and are less flexible compared to high-level fusion methodologies.

The standard approach of sensor fusion in the automotive industry is high-level, or the so-called object-level fusion or decision-level fusion. Each sensor provides high-level track information aligned in space and on a temporal basis, before they are fused by a track-level fusion module. From an economical standpoint [23], there are several advantages. At an object level, it is easy to modify the constellation on the sensor setup, for example, the incorporation of innovative new sensor types, and it is flexible in means that the interface between sensor suppliers can be defined at the object level.

However, one potential drawback is the application-specific definition of objects, leading to an extra effort in adaption. Another disadvantage is a potential loss in raw data or feature information during the object formation process. Subsection “[Motivation for Feature-Level Fusion](#)” shows a simple example comparing feature-level fusion and decision-level fusion.

Approaches like multilevel fusion and feedback fusion were investigated in the ProFusion projects [22] and are heavily dependent on the respective applications, incorporating advanced situation refinement techniques.

Grid-based fusion techniques require distance information (e.g., using stereo systems for imaging sensors) for all sensors, attended with additional

costs. Nevertheless, this contribution shows in which way it is possible to generate a grid-based representation fusing an imaging sensor with a ranging sensor by feature-level fusion.

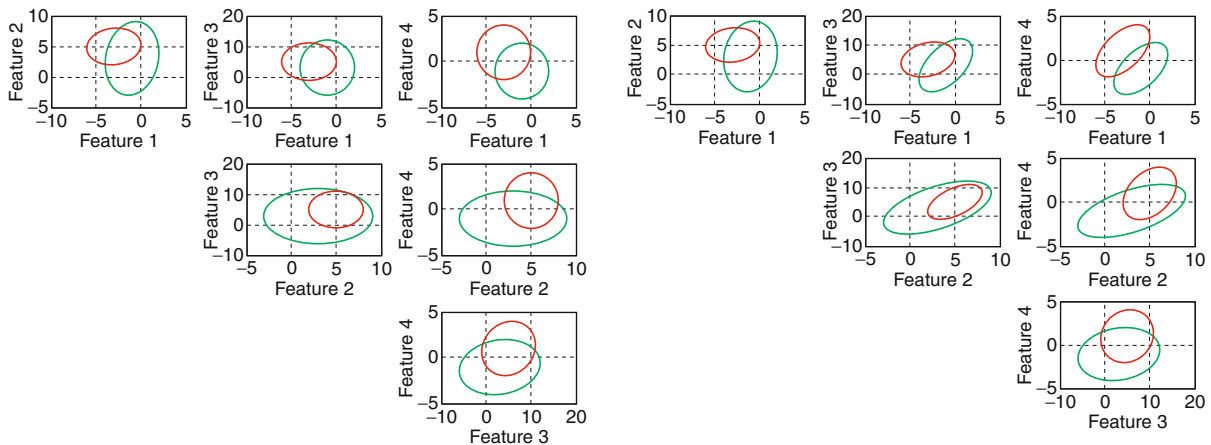
The elected fusion approach is the feature level fusion. It uses features extracted from each sensor's data building a composite feature vector to classify the objects. The major advantage of feature-level fusion is that all information, including the sensors, is combined in an optimal manner. Thus, detecting pedestrians can incorporate the information from both sensors at the same time. Even if neither of the two sensors is able to detect any single pedestrian on its own, the fusion approach at the feature level could still be capable of doing the task. Although the advantages are well known, feature-level fusion is rarely used in automotive applications. As mentioned above, this fact can be explained by unavailable low-level sensor data in automotive applications and the higher costs for the transmission with the private CAN or FlexRay link which is generally necessary.

Motivation for Feature-Level Fusion

The following example demonstrates the advantage of feature-level fusion over decision-level fusion. A two-class problem is assumed with a four-dimensional feature space extracted from the measurements of two sensors. The first sensor measures feature 1 and 2. The second sensor observes feature 3 and 4. Two classifiers have been adapted to compare a feature-level fusion approach with a decision-level fusion approach.

The feature-level approach combines all features in one Bayes classifier. The decision-level approach uses two Bayes' classifiers to separate the two classes into their two sensor subspaces of the feature space. The inferences of the two classifiers build a new feature vector, which is then classified by a trained following classifier.

For the experiment, the correlation factor between the features of different sensors is enforced simultaneously, whereas the correlation factor for features of one sensor is held constant. This is done as long as the correlation matrix persists semidefinitely. [Figure 1](#) shows the feature space of a Gaussian distribution. The features are depicted by pairs. For both classes the three-sigma bound is illustrated. The left side



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 1

Feature space of a Gaussian distribution with dimension four for a two-class problem. The features are depicted by pairs. For both classes, the three sigma standard deviation line is illustrated. The left side presents the distribution with a correlation factor 0. The right side shows it for the maximum correlation

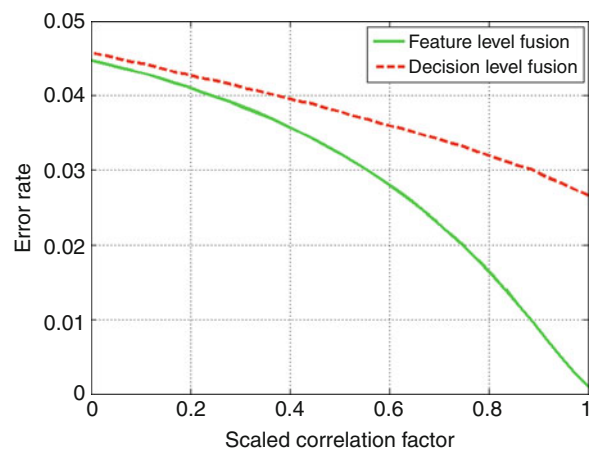
presents the distribution with a correlation factor zero between the features of different sensors. The right side shows the feature space for the maximum simulated correlation.

In Fig. 2 the achieved error rates as a function of the correlation factor are illustrated, whereby the maximum arising correlation factor is scaled to 1. Feature-level fusion clearly outperforms decision-level fusion. The performance enhancement of the feature-level fusion depends strongly on the correlation factor between the features of the two sensors. The risk of applying feature level fusion in place of decision-level fusion is the increased feature space for the classifier. With the expansion of the feature space, the risk of over-adaption in the training stage increases.

System Overview

Figure 3 shows an overview of the structure and the relations between the information sources and the implemented subsystems necessary to construct a pedestrian warning application. The application is composed of three subsystems.

The first part combines the information provided by the camera sensor and the radar sensor to detect pedestrians reliably and robustly based on the presented feature level fusion approach.



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 2

Comparison of error rate between feature- and decision-level fusion: The performance enhancement of the feature level fusion depends on the correlation factor between the features of the two sensors

The second part estimates the road course using the information provided by a digital map (i.e., a fixed road shape in front of the vehicle), the radar observations, the inertial vehicle measurements, and the initial position of a GPS receiver. The goal is to fit the road shape

into the radar measurements, which is achieved by testing different global vehicle positions in a local area around the last estimated vehicle position. The GPS receiver provides the first position estimate to initialize the algorithm.

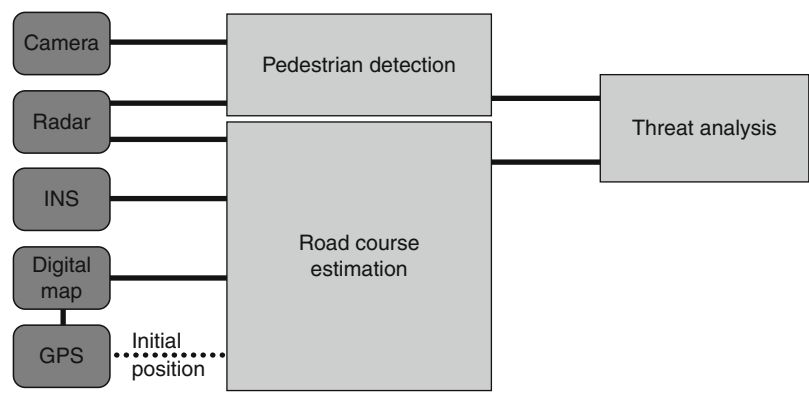
The third part combines the results of the previous subsystems to analyze the relevance of each detected pedestrian depending on their position on the road.

This threat analysis gives the opportunity to direct the driver's attention to a situation of imminent danger.

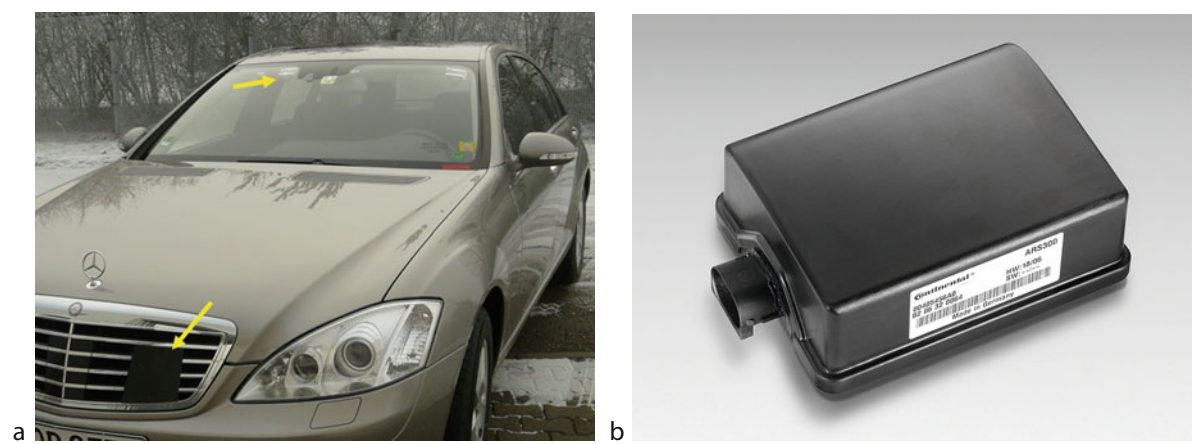
Sensor Configuration

The presented pedestrian recognition system fuses data from a camera sensor and a radar sensor. Figure 4a shows the sensor mounting positions in the research vehicle: the camera behind the windshield and the radar sensor in front of the radiator grill.

The vision component corresponds to the Mercedes-Benz Night View Assist, consisting of an NIR CMOS camera sensor and infrared headlamps providing roughly the same range of visibility as high beam



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 3
System overview: The application is composed of three subsystems, the pedestrian detection, the road course estimation, and threat analysis



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 4
(a) Research vehicle with sensor mounting positions and (b) 3rd generation ACC advanced radar sensor ARS300 from the Continental AG

headlamps – but without impairing the vision of the oncoming traffic. The camera is configured with 12 mm optics and supplies 12 bit gray-scale VGA images.

The advanced radar sensor ARS300 is a new mechanically scanning 77 GHz FMCW radar sensor developed by the Continental AG. Figure 5 gives an impression of the radar measurement raw data which consists of the intensity $i(r, \varphi)$ of the reflected signal in dB and the relative speed $v_{\text{rel}}(r, \varphi)$ of reflecting objects in kph with range $\{r \in \mathbb{Z} | 0\text{m} \leq r \leq 199\text{m}\}$ and azimuthal angle $\{\varphi \in \mathbb{Z} | -8^\circ \leq \varphi \leq 8^\circ\}$. In this configuration each cell (r, φ) has the dimension of $1\text{ m} \times 1^\circ$. The resolution in relative speed is specified with 2.76 kph.

Sensor Alignment

In order to associate measurements from multiple sensors, the alignment of each sensor to the vehicle system (DIN70000) needs to be accurately determined. The radar system is calibrated using the alignment procedure developed by the Continental AG. The centered radar beam is aligned in azimuth and elevation to the longitudinal vehicle axis. However, to calibrate the camera system, the fundamental matrix $\mathbf{F} = \mathbf{F}_{\text{intr}}\mathbf{F}_{\text{extr}}$ containing the intrinsic parameters and the extrinsic parameters needs to be estimated. While the intrinsic camera parameters (e.g., focal length, principal point, skew and distortion coefficients) are estimated using a standard camera calibration

technique [24–26], the extrinsic parameters (mounting position and Euler angles) are assessed by solving a nonlinear least-square data fitting problem of the calibration target for N positions:

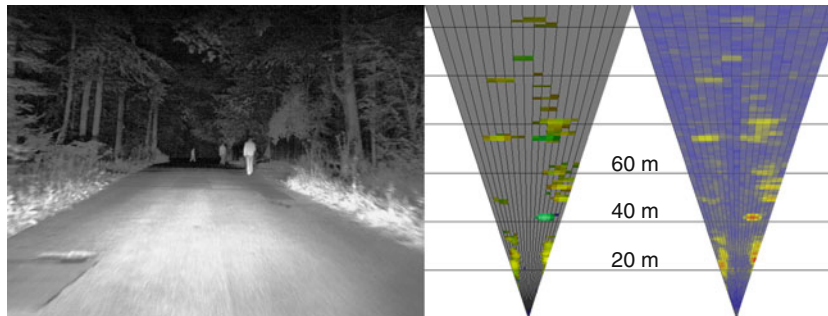
$$\hat{\mathbf{F}}_{\text{extr}} = \arg \min_{\mathbf{F}_{\text{extr}}} \sum_{i=1}^N \left(\mathbf{F}_{\text{intr}} \mathbf{F}_{\text{extr}} \mathbf{x}(r_i, \varphi_i) - \begin{pmatrix} u \\ v \end{pmatrix}_i \right)^2,$$

where $\mathbf{x}(r_i, \varphi_i)$ and $(u, v)^T$ determine the position of the calibration target in the vehicle coordinate system and in the image plane, respectively. An optimal calibration target in the radar plane is a trihedral corner reflector [27], reflecting the emitted electromagnetic wave directly backward to the radar sensor. A 2D Gaussian function is fitted to the intensity data $i(r, \varphi)$ to improve the measured discrete position of the target in the radar frame with $\Delta r = 1\text{ m}$ and $\Delta \varphi = 1^\circ$. The tip $(u, v)^T$ of each corner position is used as a reference point in the image plane. Figure 6 shows the result of the calibration.

Boosted Feature-Level Sensor Fusion

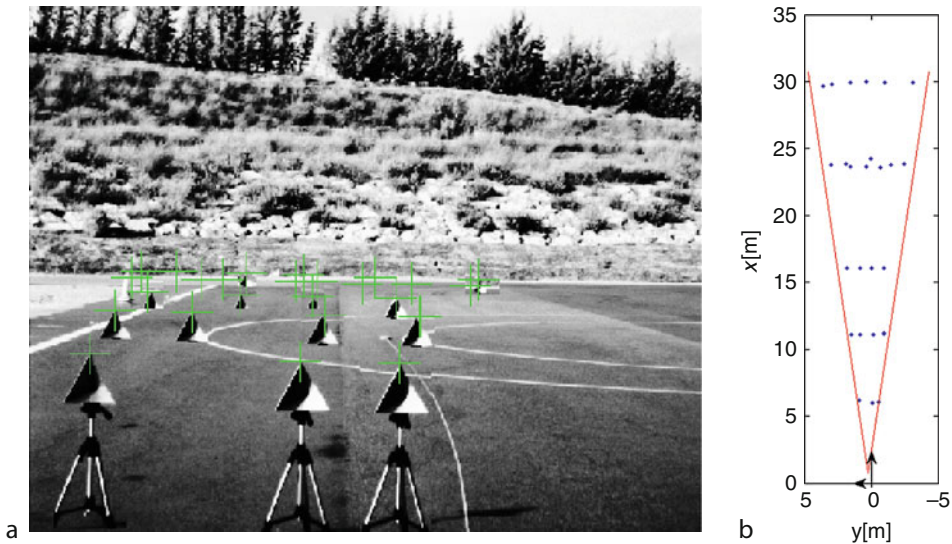
Fusion Architecture

In order to increase robustness of the pedestrian detection, the system combines the information from complementary sensors (i.e., a camera and a radar sensor) on a feature level using the supervised training algorithm AdaBoost [28]. The sensor fusion architecture consists of three layers (see Fig. 7):

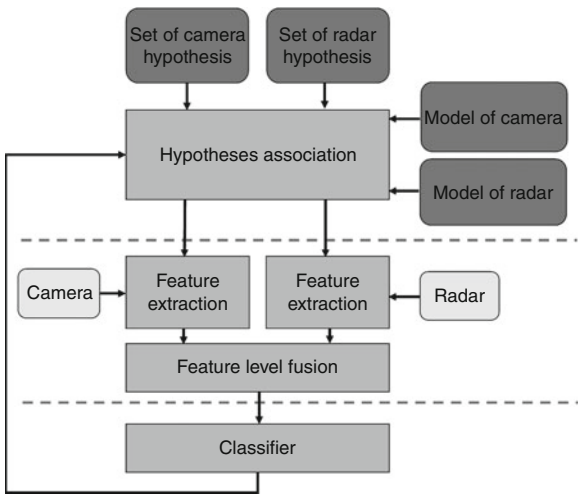


Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 5

Radar measurements: The left image depicts the experienced scene from the camera. The image in the middle shows the radar intensity (value) and relative speed (hue). The three moving pedestrians can be recognized by the green hue component at about 41 m, 75 m, and 108 m. The right image shows all radar intensity values (blue corresponds to low and red to higher intensity)



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 6
Calibration result: (a) Composed image of M target positions with $M \subseteq N$ and the projection of N target positions into the image. (b) N radar measurements of the trihedral corner reflector in the vehicle coordinate system



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 7
Sensor fusion architecture

1. Firstly, hypotheses are generated within the search space. Hypotheses originating from the different sensors are associated based on model information (camera and radar), and eventually on detection results from the previous detection stage

(coarse-to-fine search, see section “[Multi-Sensor Hypotheses](#)”).

2. In the second processing step, the features are calculated and organized in a combined set for both sensors.
3. The classification itself is done in the third processing step based on the calculated features.

The layers two and three are optimized simultaneously. Feature level fusion is achieved in the following way: Features are obtained from each sensor stream simultaneously. These features are held in one set. During the training stage, the source of the individual features and their type do not matter. AdaBoost selects incrementally those features having the lowest weighted error on the training set, until the algorithm achieves a predefined correct and false detection rate. This technique assures an optimal selection and weighting of the features from both sensors for the classification task.

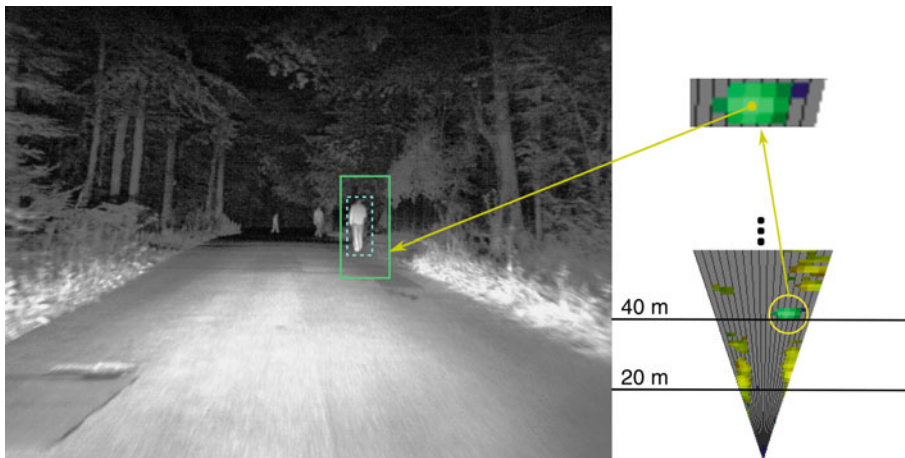
All features are defined within search windows belonging to the different sensors. The tuple of search windows (one for each sensor input) describes a hypothesis presented to the classifier. The latter assumes that both search windows identify the same localization of the scene. This requires an accurate mapping between the camera plane and the radar

plane. However, combining multiple sensors also increases the input space and, consequently, increases the computational effort due to the ambiguous mapping between the camera plane and the radar plane.

The radar measurement $i(r, \varphi)$ defines an ROI in the camera image (see Fig. 8), which is much bigger than the projected object at the measured position (r, φ) . Moreover the object height cannot be measured by the radar system. Thus multiple search windows in the image are possible to be associated to a single radar

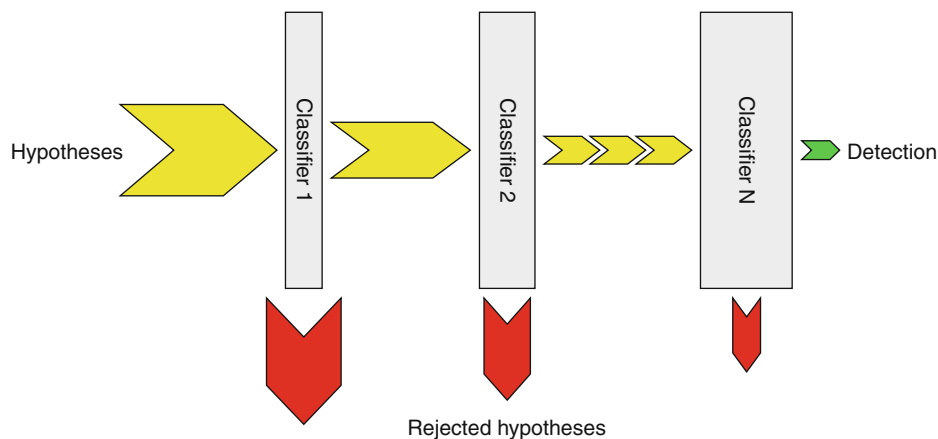
measurement $i(r, \varphi)$. Section “Multi-Sensor Hypotheses” describes the association rule in detail.

In order to realize a real-time detection system, the presented fusion system incorporates cascaded classifiers introduced by Viola and Jones [29], [30]. Such a classifier is capable of separating regions in the input space into two classes, namely, the object and background class using a chain of classifier stages (see Fig. 9). With an increasing number of stages, the complexity of these classifiers rises by the number of



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 8

Ambiguous mapping between the camera and the radar plane: A radar measurement $i(r, \varphi)$ defines a much bigger ROI (green) in the image plane than the corresponding pedestrian object (dashed blue). Thus the measurement (r, φ) belongs to multiple search windows in the corresponding image area



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 9

Cascaded classifier: Each stage discards hypotheses and passes promising samples onto the successive and more complex stage

features used. Each classifier stage is trained using the AdaBoost algorithm to choose the most significant features from both sensors. Depending on the discriminative power of the features from the camera sensor and the radar sensor, the resulting cascade can perform the classification task very efficiently. Subsection “Radar Features” describes the derived radar features in detail.

The AdaBoost Approach to Machine Learning

Boosted classifiers and the related AdaBoost algorithm introduced by Freund and Schapire [28] are very well described in [31–33]. The basic idea is that a set of weak classifiers build a strong classifier. The decision of the strong classifier is a weighted voting of the decisions of the weak classifiers. The weak classifiers are selected subsequently in the trainings phase. On each round, the weights of incorrectly classified examples increase, the weights of correctly classified examples decrease, so that the next weak classifier focuses more on incorrectly classified samples. The importance of the weak classifier depends on its error on the weighted sample set.

A pseudo-code description of the multi-sensor algorithm regarding a two-class classification scenario is given in Fig. 10.

The algorithm requires a multi-sensor training set of N samples with $(x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(M)}, y_1), \dots, (x_N^{(1)}, x_N^{(2)}, \dots, x_N^{(M)}, y_N)$ as input where $x_j^{(k)}$ represents the input of the k th sensor and $y_j \in \{-1, +1\}$ the corresponding manually extracted label (i.e., -1 for a background and $+1$ for an object sample). For all iterations $t = 1, \dots, T$ AdaBoost calls a weak learning algorithm to train the base classifier with importance $\alpha_t \in \mathbb{R}$ on a weighted version of the training set by minimizing the weighted error ε_t :

$$\varepsilon_t = \min_{i,k} \sum_{j=1}^N w_{i,j} \left| \frac{h_{t,i}^{(k)}(x_j^{(k)}) - y_j}{2} \right|$$

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right).$$

Thus each round AdaBoost chooses the best weak learner $h_i^{(k)}$ from the set of specified features of all

Require:

- N samples $(x_j^{(1)}, x_j^{(2)}, \dots, x_j^{(M)}, y_j)$ for M sensors with $j = 1, \dots, N$ and $y_j \in \{-1, +1\}$
- τ : minimum detection rate
- T : maximum number of selectable weak classifiers

Ensure:

The final strong learner decision $H(x)$ is composed of the most discriminative weak learners $h_t^{(k)}$ and achieves the required detection rate.

- 1: initialize $w_j = \frac{1}{N}$
- 2: **for** feature $t = 1, \dots, T$ **do**
- 3: determine weak learner $h_t^{(k)}$ that minimizes the weighted error ε_t on the training data
- 4: choose α_t , the importance of $h_t^{(k)}$
- 5: update sample weights and normalize so that w remain a distribution:

$$w_j = \frac{1}{n_t} w_j \exp^{-\alpha_t y_t^{(k)}(x_j^{(k)})}$$

with n_t as a normalizing factor

6: **end for**

- 7: adjust θ of the final strong learner to realize the minimum detection rate τ :

$$H(x) = \text{sgn} \left(\left(\sum_{t=1}^T \alpha_t h_t^{(k)}(x^{(k)}) \right) - \theta \right)$$

Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 10

Pseudo-code description of the multi-sensor AdaBoost algorithm

sensors (e.g., an image feature or a radar feature). The final strong learner decision H is a weighted majority vote of the chosen T weak learners:

$$H(x) = \begin{cases} +1 & : \sum_{t=1}^T \alpha_t h_t^{(k)}(x^{(k)}) \geq 0 \\ -1 & : \text{else.} \end{cases}$$

Applying this algorithm to each stage of a cascaded classifier, the decision in the first layers with less weak learners can be dominated by one of the sensors with the most discriminative power of their features.

Image Features

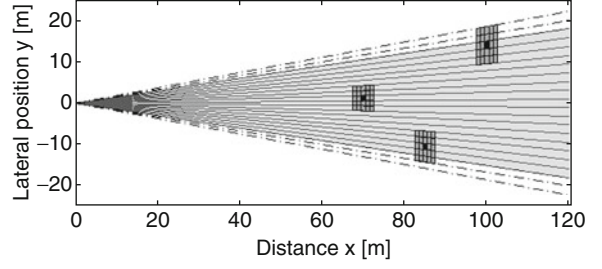
In the image plane, an object (e.g., pedestrian) is represented by a rectangular search window, which is defined by its upper left and lower right corner with a fixed width-to-height ratio (e.g., 0.5). The feature set used in the image plane consists of Haar-wavelet filters [34, 35], which vary in size and position within the rectangular search window. These features have been successfully applied in many pedestrian recognition systems [16, 17, 35], and are able to compute efficiently using integral images.

Radar Features

In the radar plane, the feature set is composed of two different filter subsets:

- Subset S_1 consists of filters describing the average value of the reflected object intensity
- Subset S_2 consists of complex signal filters with a structure similar to the radar CFAR (constant false alarm rate) filters [27]

Contrary to the image plane, a pedestrian cannot be distinguished clearly in the radar plane because the radar signal of a typical pedestrian is similar to many other objects, for example, posts, street lamps, and trees. In other words, these filters are not capable of unambiguously classifying a pedestrian. However, they can be used to reject background samples very efficiently. The manually labeled pedestrian in the radar image at position (r, φ) defines the center of the search window with a fixed size $5 \text{ m} \times 5^\circ$. Figure 11 shows the expanded field of view by mirroring the corresponding beam directions at the left and the right border. This allows for apply the filters to all beam directions φ .



Night Vision Pedestrian Warning in Intelligent Vehicles.
Figure 11

Definition of the search window in the radar plane projected onto the Cartesian vehicle plane: Each position (r, φ) indicates the centre of the search window with a fixed size of $5 \text{ m} \times 5^\circ$. The field of view is expanded in φ to apply the filters to all beam directions

1. *Average object intensity filter*: The reflected signal intensity $i(r, \varphi)$ can be used to very efficiently separate between object and background. This can also be seen in Fig. 5 (in the radar intensity measurements). Within the search window, a rectangular filter is used to calculate the mean signal intensity. Thus subset S_1 consists of different rectangular filters varying in position and size within the search window.
2. *Complex signal filter*: The complex signal filter combines the reflected signal intensity $i(r, \varphi)$ and the relative object velocity $v_{\text{rel}}(r, \varphi)$ to detect object clusters of similar velocity. With regard to the theory of radar systems, the measured object intensity and relative velocity represent the highest peak in the Doppler spectrum at the measured position (r, φ) . Hence the complex number $z(r, \varphi)$ describes the radar measurement (i, v_{rel}) at position (r, φ) equally:

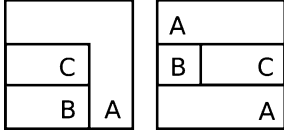
$$z(r, \varphi) = A_z \exp^{j\pi\omega}, z \in \mathbb{C}$$

$$A_z = i(r, \varphi)$$

$$\omega = f(v_{\text{rel}}(r, \varphi)) + v_{\text{ego}}, \{\omega \in \mathbb{R} \mid |\omega| \leq 1.0\},$$

where A_z corresponds to the measured intensity, v_{ego} defines the driving speed, and ω describes the measured absolute velocity of measured objects after the normalization.

A complex signal filter is composed of three rectangular zones (see Fig. 12): a background zone A, a blind



Night Vision Pedestrian Warning in Intelligent Vehicles.

Figure 12

A complex signal filter is defined by three zones:
a background zone A, a blind zone B and a test zone C with
 $C \subseteq B \subseteq A$

zone B, and a test zone C with $C \subseteq B \subseteq A$. This corresponds to the CFAR filter structure [27] extended to two dimensions. Arranging the three zones in all positions and sizes within a search window of the size $5 \text{ m} \times 5^\circ$ results in 44,100 different filters. The real-valued filter value is defined by:

$$\left| \frac{|A \setminus B|}{|C|} \sum_{\forall(r, \varphi) \in C} z(r, \varphi) - \sum_{\forall(r, \varphi) \in A \setminus B} z(r, \varphi) \right|$$

which is equivalent to the norm of normalized differences between the integrated complex values $z(r, \varphi)$ over C and $A \setminus B$. Figure 13 shows the filter calculation and the resulting filter value which describes a local similarity measure of velocity differences. These filters suppress noisy measurements indicated by a small complex modulus and a random complex argument in $z(r, \varphi)$. However, the defined filters cannot distinguish between moving pedestrians and small stationary objects with a similar intensity pattern (e.g., posts): both result in high filter responses. In contrast, low filter responses are caused by bigger objects with intensity patterns different from pedestrian patterns (e.g., cars, houses).

In order to reduce the computational complexity, the algorithm decomposes the complex number z into their components $\text{Re}\{z\}$ and $\text{Im}\{z\}$ and uses integral images to calculate the filter value components independently.

Multi-Sensor Hypotheses

The presented cascaded classifier requires a tuple of search windows identifying the same scene in the image plane and in the radar plane. Search windows are generated in a range from 20 to 120 m, which corresponds to pedestrian sizes in the image plane from 120 to 21 px. Due to the ambiguous mapping

between both sensors (see Fig. 8), this results in multiple search windows in the image plane associated to one search window in the radar plane. In order to efficiently perform on the increased input space, the classifier scans the image plane restricted to the ROI using the relaxed *flat world assumption* presented in [19] that compensates the pitch of the vehicle.

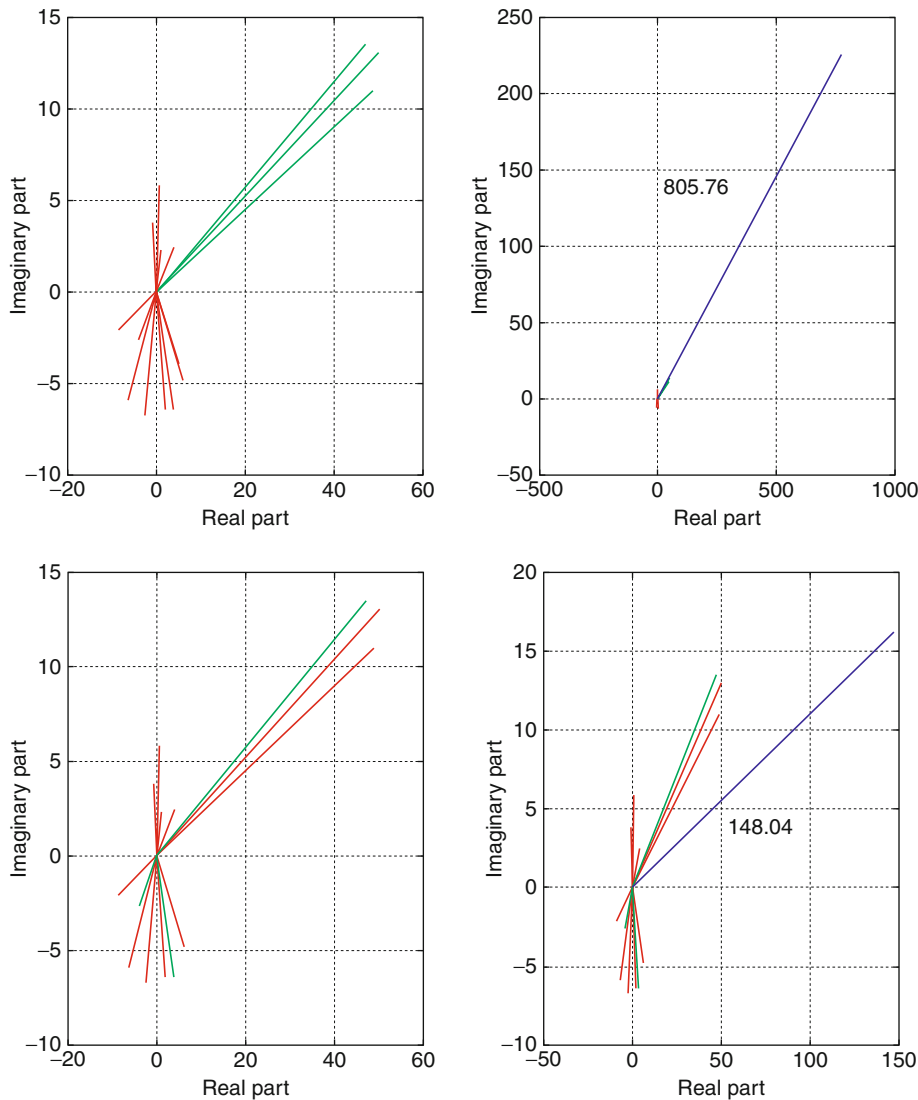
Using the camera model, the search window height of a minimum size and a maximum size pedestrian can be calculated at each distance r of the radar sensor. These heights represent the limits of scaled versions of the minimum search window height at distance r . The combination of the restricted image plane and the set of search window heights define the resulting grid of search windows in the image plane associated to the search windows with distance r in the radar plane.

The resulting amount of hypotheses to be classified remains unmanageable for real-time applications. With a coarse-to-fine search, generating a set of search window grids in the image plane with different position sampling increments the problem can be solved. First, the detector classifies the coarse grid and continues the classification in the fine grid for positive classified search windows.

Global Positioning and Road Course Estimation

The previous sections presented the feature level sensor fusion approach to detect pedestrians at night using an NIR camera and an imaging radar. In order to realize a warning system which directs the driver's attention primarily to the relevant objects, the information about the pedestrian's position is combined with a new road course estimation approach. Instead of estimating the shape of the road, this contribution estimates the position on a digital map and extracts the preceding road course directly from the map. In contrast to loosely and tightly coupled INS/GPS solutions, the main concept of this contribution is based on the method to localize a robot in a known defined environment, also known as Monte Carlo Localization (MCL) [36, 37].

MCL algorithm uses a particle filter approach [38]. Each particle describes a position and an orientation of the robot in a restricted environment which has been mapped using the sensor measurements of the robot (e.g., ultrasonic, radar, laser, and camera). At the beginning, the position of the robot is unknown. Thus each position in the restricted environment is possible,



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 13

Complex signal filter response: Given a search window, the corresponding radar measurements (i, v_{rel}) are assigned to the filter zones (left figures): background (red), blind (the measurements which are not taken into account) and test (green). The upper graphics show the optimal filter and the lower graphics show a suboptimal filter in the complex plane related to the given radar measurements. The filter response (right graphic) is shown by the complex number in blue with the corresponding real-valued filter result as text

leading to a uniform distribution of the particles on the map. In order to localize the robot on the map, the algorithm verifies each position based on the local sensor measurements. After a few iterations, the particles will converge to the most likely position and thus the robot is localized.

A significant difference to a robot is that the environment of a passenger car corresponds to an unrestricted network of roads. Thus, the position cannot be uniformly distributed in the initialization cycle. This problem can be solved by means of a low-cost GPS sensor which restricts the area of possible initial

positions to the maximum position uncertainty of ± 25 m. The position and the orientation of the ego vehicle are described in UTM (Universal Transverse Mercator) coordinates [39] which project meridian segments into a rectangular plane with negligible distortion error. In order to verify a position hypothesis, the observations of the imaging radar are compared with the extracted road course of an electronic horizon component (e.g., [40]). Note that the electronic horizon component also performs a map matching using the GPS position and a digital map. But the calculated position does not regard the width of the road, and the orientation of the vehicle cannot be estimated precisely. The resulting state vector of a position hypothesis is defined by

$$\underline{\mathbf{q}}_k^{(i)} = (e_k^W, n_k^W, \Phi_k^W, b_k^W)^{(i)}.$$

The scalar components $e_k^W, n_k^W, \Phi_k^W, b_k^W$ of the particle number (i) denote the position in UTM (east and north coordinate), the orientation of the vehicle, and the width of the road for iteration k .

The main steps of the positioning algorithm using a particle filter approach can be outlined as follows:

1. Sample from the probability density $p(\underline{\mathbf{q}}_{k-1} | \underline{\mathbf{y}}_{0:k-1})$ of the previous iteration generates a new set of particles. Each particle $\underline{\mathbf{q}}_k^{(i)}$ describes a vehicle coordinate system in a global UTM coordinate system.
2. Apply a deterministic and a stochastic drift to predict the hypotheses to the most likely position of the current iteration. The deterministic portion corresponds to the egomotion of the ego vehicle during an iteration step which is estimated by an extended Kalman filter (EKF). The stochastic portion models the process uncertainties by an unbiased Gaussian and yields to a random distribution of equally sampled particles.
3. Calculate the likelihood for each particle based on the measurements of the imaging radar sensor. This step is based on an intensity model of the road, which expects that the road itself is free of objects and narrowed by high intensities from stationary objects. Hence a preprocessing step of the radar measurements is required to compensate the high intensities caused by moving objects on the road.
4. The final step is to generate the probability density $p(\underline{\mathbf{q}}_k | \underline{\mathbf{y}}_{0:k})$ based on the weighted particle set.

The following subsections discuss the egomotion estimation, the radar preprocessing, and the likelihood function in detail.

Egomotion Estimation

In order to measure the egomotion of the vehicle, an EKF approach is used which considers the inertial sensor measurements providing information about the turn rate $\dot{\Psi}$ and the wheel rates $d_{fl,fr,rl,rr}$ of all four wheels. As discussed in [41], a *Constant Turn Rate and Velocity* (CTRV) motion model is totally sufficient to describe the common movements of a vehicle. The state space and the continuous differential equations are defined by

$$\begin{aligned} \underline{\mathbf{x}} &= [x, y, v, \Psi, \dot{\Psi}]^T \quad \text{with} \\ \dot{x} &= v(t) \cos \Psi(t), \\ \dot{y} &= v(t) \sin \Psi(t), \\ \dot{v} &= 0, \\ \ddot{\Psi} &= 0. \end{aligned}$$

The scalar components x, y, v, Ψ and $\dot{\Psi}$ denote the longitudinal and the lateral (to the left) movement, the longitudinal velocity, the rotation angle, and the turn rate. This curvilinear motion model involves a nonlinear state transition as well as a nonlinear measurement model using a four-wheel vehicle model. The discretized formulation based on the sample time T can be derived by solving

$$\underline{\mathbf{x}}_{k+1} = \underline{\mathbf{x}}_k + \int_t^{t+T} \underline{\mathbf{x}}(\tau) d\tau.$$

This yields to

$$\begin{aligned} x_{k+1} &= x_k + \frac{v_k}{w_k} (-\sin(\Psi_k) + \sin(w_k T + \Psi_k)), \\ y_{k+1} &= y_k + \frac{v_k}{w_k} (\cos(\Psi_k) - \cos(w_k T + \Psi_k)), \\ v_{k+1} &= v_k, \\ \Psi_{k+1} &= \Psi_k + w_k T, \\ w_{k+1} &= w_k \quad (\text{the discrete turn rate}), \end{aligned}$$

when $|w_k| \rightarrow 0$ then use

$$\begin{aligned} x_{k+1} &= x_k + v_k T \cos(\Psi_k), \\ y_{k+1} &= y_k + v_k T \sin(\Psi_k). \end{aligned}$$

In order to handle the asynchronous arriving turn rate and wheel rate measurements, the EKF implementation corresponds to the sequential measurement update formulation presented in [42], which is further proved to be efficient and numerically stable. The filter consistency is maintained by testing each measurement innovation against its predicted innovation uncertainty in a three-sigma bound which covers 99.7% of the expected measurements. A failed test will reset the filter to the initial state covariance, which happens when the CTRV process model does not represent the vehicle's motion (occurs mostly in sharp turns and when the vehicle accelerates). The accuracy of the egomotion estimation algorithm is shown in Fig. 14.

The required incremental movement between consecutive radar measurements is provided by a reset operation after each cycle of the positioning algorithm. For this purpose, the scalar components x , y , and Ψ are set to zero and the corresponding entries in the covariance matrix are set to their initial values.

Radar Preprocessing

Moving obstacles in the field of view of the radar sensor have a huge impact on the evaluation of position hypotheses because the algorithm assumes a road model with low intensities on the road and high intensities on the road border. Due to the high radar cross section of vehicles, a moving vehicle on the road

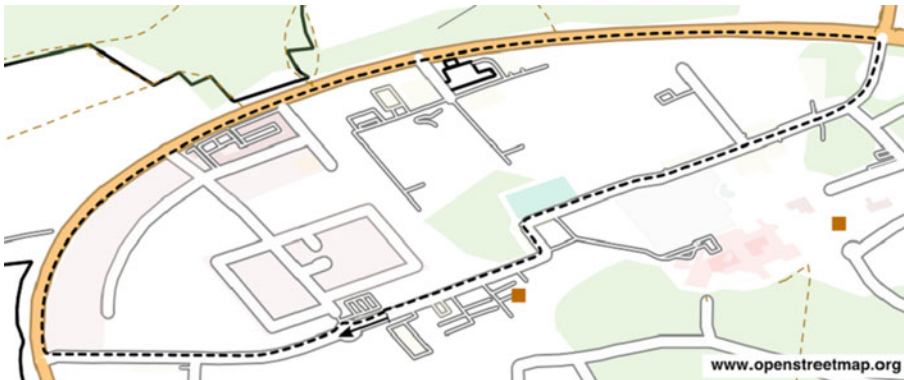
violates this assumption. In order to diminish this impact, a preprocessing step is applied replacing the intensity measurements by their corresponding noise level value (measured by the radar sensor for each cell (r, φ)) when the measurement represents a moving object. The moving object decision is based on the measured radial relative velocity v_D^R (Doppler), its direction φ^R and the estimated velocity of the ego vehicle v_{ego} .

$$\text{Moving object} \Leftrightarrow |v_D^R \cos(\varphi^R) - v_{ego}| \geq \gamma_v,$$

where γ_v denotes the velocity threshold.

Evaluation of Position Hypotheses

Each hypothesis or particle defines a position and an orientation of the vehicle in UTM coordinates. Given that position and the road course from the digital map, an individual course of the preceding road segment can be described in vehicle coordinates. This provides the possibility to establish a relationship between the digital map data and the radar observations of the environment. The estimation of the road width is used to define two polylines describing the boundary of the road. The next step transforms these two polylines for each particle into the bilinear interpolated Cartesian representation of the preprocessed polar radar intensity measurements (see Fig. 15). Thus, a goodness of fit measure can be computed for each particle. This evaluation is based on



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 14

Egomotion accuracy analysis: The figure shows the estimated course of a closed-loop scenario of length 4.5 km in UTM coordinates. The resulting estimation error between the starting and the end position (represented by the arrow in driving direction) is about 2.4 m



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 15

Results of the position estimation process: The small maps show the particle distribution at initialization of the algorithm after the 2nd and after the 5th iteration. The sensor observations of the camera and the radar sensor as well as the estimated road course after the 5th iteration step is shown above

the mean intensity value between the polylines \bar{Y}_1 and the mean intensity value below the polylines \bar{Y}_2 according to the logistic function

$$g(\bar{Y}_2 - \bar{Y}_1) = \beta (1 + \exp(-(\bar{Y}_2 - \bar{Y}_1 - \alpha)))^{-1}.$$

The inflection point α denotes the barrier to evaluate a hypothesis as a positive road course candidate. The nonlinear characteristic of $g(\cdot)$ improves the convergence behavior. Furthermore, the factor β models the ratio of the estimated road course in the radar's field of view.

Figure 15 shows three iteration steps of the positioning algorithm. The expected road course after the 5th iteration is shown in the Cartesian representation of the radar measurements. Both polylines are partly outside of the field of view which yields to the ratio $\beta < 1.0$.

Threat Analysis

After the detailed description of the pedestrian detection system and the road course estimation system, the

final task is to fuse both information sources to realize a pedestrian warning system. A range-dependent object warning is straightforward using the position information provided by the detection system. The more challenging task, discussed here, is to warn the driver about objects depending on their lateral position to the road which requires a threat analysis based on both information sources. Figure 16 shows the desired warning scheme with low ratings on the sidewalk and the highest rating in the center of the right lane. Moreover, a pedestrian on the right sidewalk poses a higher threat to the driver than a pedestrian on the left sidewalk, under the assumption of driving in the right lane.

An advantage of the presented system architecture is the shared usage of the radar data in the detection system and in the road course estimation system. This gives the opportunity to measure directly the relevance of an object. The detection system describes each object position in polar radar coordinates (r_i, φ_j) and in image coordinates (u, v, w, h) . The road course estimation system delivers the polylines of the road boundary

in Cartesian radar coordinates which yields to a lateral resolution problem when using (r_i, φ_j) directly. Hence the image position is used to define a more precise object position in (r_i, φ_j) . The calculation is based on a linear interpolation between the calibrated left and right image columns of the object beam φ_j .

This refined position defined by (r_i, φ_j^*) is then transformed into the Cartesian view to measure the distance d to the left polyline of the estimated road course. The final warning level of a detected object corresponds to

$$\eta(d^*) = \exp\left(-\frac{(d^* - 0.75)^2}{2\sigma^2}\right).$$

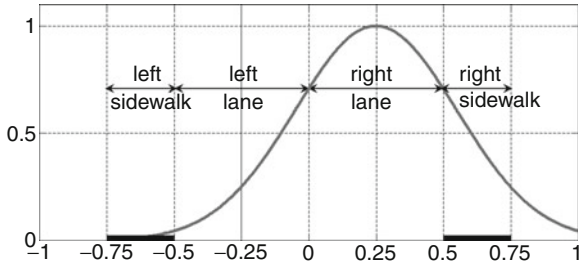
The distance d^* is normalized by the estimated road width and σ^2 chosen to 0.1 defines the warning focus. Figure 17 shows the result of the realized threat analysis. The warning level changes from red to green for a pedestrian walking from the center of right lane to the left sidewalk.

Experiments

Pedestrian Detection System

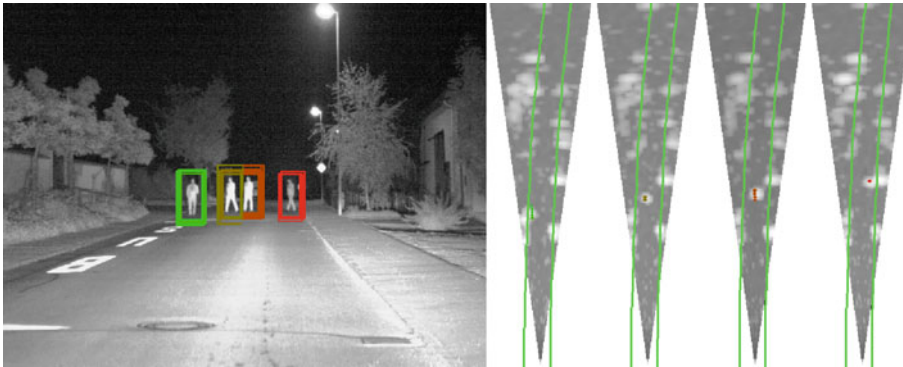
The following subsections discuss the performance of the feature-level fusion approach used in the pedestrian detection system. This requires methods for ground truth generation, a set of recorded sequences for training and testing and an evaluation criterion.

Defining a Ground Truth Each supervised classification task requires a consistent ground truth for training and evaluating a classifier. With regard to a multi-sensor configuration, consistency is much more important because an object label in one sensor plane has to identify the same object in the other sensor plane. Figure 17 shows a typical scene of pedestrians in a suburban environment. Labeling the pedestrians in



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 16

The desired warning scheme on a two-lane road: The most threatening object position is in the centre of the right lane. The width of the road is normalized to one



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 17

Result of the threat analysis: The left image shows a walking pedestrian from the right lane centre at iteration k to the left sidewalk at iteration $k + N$. While the pedestrian is walking to the left sidewalk the warning level of the pedestrian is decreasing from red to green. The figures on the right show the corresponding pedestrian position as well as the estimated road course in the Cartesian radar measurement view

the image is rather simple and easy to learn. However, labeling the same pedestrian in the radar image is much more complicated and requires an expert or some kind of label assistance. Different label assistance options are possible:

1. Monocular distance estimation assuming a fixed pedestrian size of the label in the image plane
2. CFAR detector applied to the projected epipolar line in the radar plane given by an assumed minimum and maximum pedestrian size
3. Stereo distance estimation analyzing the pixel disparity within the label in the image plane

Note that they all utilize the information provided by the pedestrian label in the image plane. While option 1 fails due to poor label accuracy, especially in a high detection range, and option 2 fails due to the fixed structure of the CFAR filter mismatching the pedestrian with other objects on the epipolar line, option 3 achieves the most promising results. Using this assistance the labeler can use the initial conjecture and correct the position if necessary. This method implies also a stereo camera setup with a large baseline in order to achieve good distance estimates of labeled pedestrians up to 120 m.

Training and Evaluation Settings To demonstrate the performance of the presented detection system, two pedestrian detectors were trained: an image-based reference detector using only the Haar-wavelet filters in the image plane and a fusion detector using additionally the defined radar features.

The training set consists of 36,562 images taken from 122 pedestrian sequences and 71 non-pedestrian sequences recorded nightly in the period from July 2007 to June 2008. The collection of images contains 17,833 pedestrian samples and represents a good mixture of pedestrians in rural and suburban environments within 120 m. Both detectors use 30,000 background samples selected randomly in each stage of the cascade from the image set.

The test set contains 7,887 pedestrian samples in 8,449 images. In order to compare the performance of both detectors equitably, an evaluation algorithm analyzes the detections in the image plane restricted by the field of view of the radar sensor. All detections are separated in true and false positives using a coverage

measure $\text{cov}(A,B)$ between the rectangular detection A and the pedestrian label B:

$$\text{cov}(A, B) = \frac{A \cap B}{A \cup B}.$$

The threshold γ_{\min} defines the minimum required coverage to assign detections to the set of true positives.

Detection Results The trained fusion detector significantly outperforms the reference detector while requiring less stages and thus less features (see Table 1). Furthermore, AdaBoost selects up to 38% of the features from the set of radar features which emphasizes their discriminative power. These radar features afford a fast and cost-effective reduction of hypotheses in early stages of the cascaded fusion detector. Figure 18 compares the computational cost of both classifiers in environments of different complexity. Figure 19 shows the ROC curves comparing the performance of the fusion detector and the image based reference detector for $\gamma_{\min} = \{0.5; 0.6\}$. For the fusion detector, the detection rate is higher and the false alarm rate is lower than the performance values of the reference system using only the information provided by the NIR camera. Furthermore, system performance increases with the required minimum coverage γ_{\min} . This behavior indicates the detection quality in spatial resolution: While the fusion detector uses the information provided by the radar sensor to suppress invalid hypotheses and effectively achieves a grid based fusion approach, the reference classifier cannot distinguish between pedestrian hypotheses at different distances. Hence, the latter produces detection results incapable of localizing the pedestrian in space. Figure 20 shows the precise pedestrian localization of the fusion system compared to the uncertainty of the reference detector.

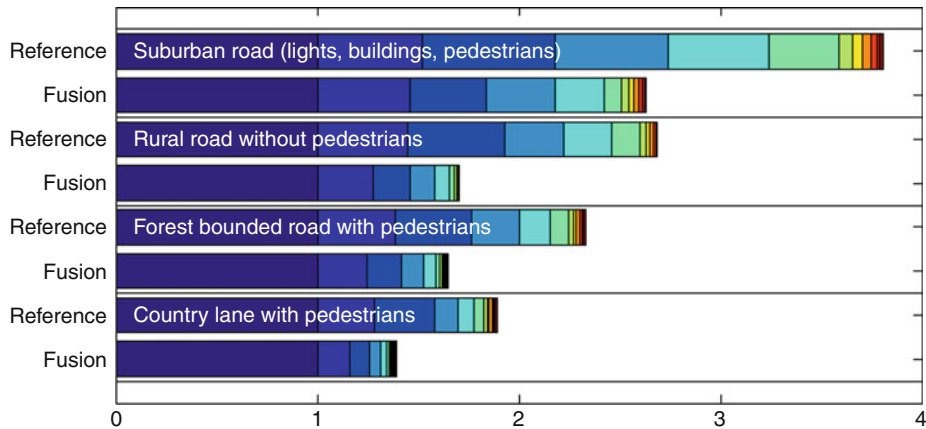
Road Course Estimation

In order to evaluate the estimated road course, or rather the estimated vehicle position, a ground truth is required.

Night Vision Pedestrian Warning in Intelligent Vehicles.

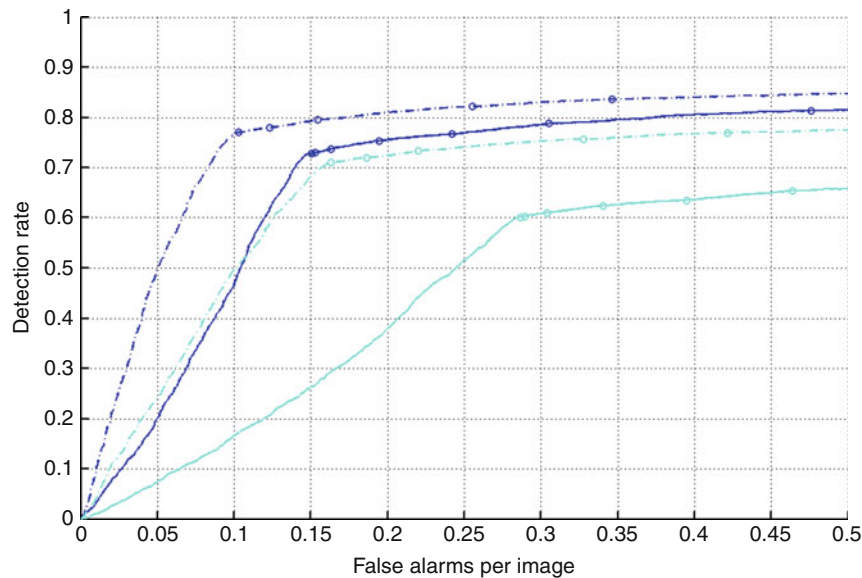
Table 1 Detector characteristics

	Stages	Image features	Radar features
Fusion	28	1061	407
Reference	30	1621	–



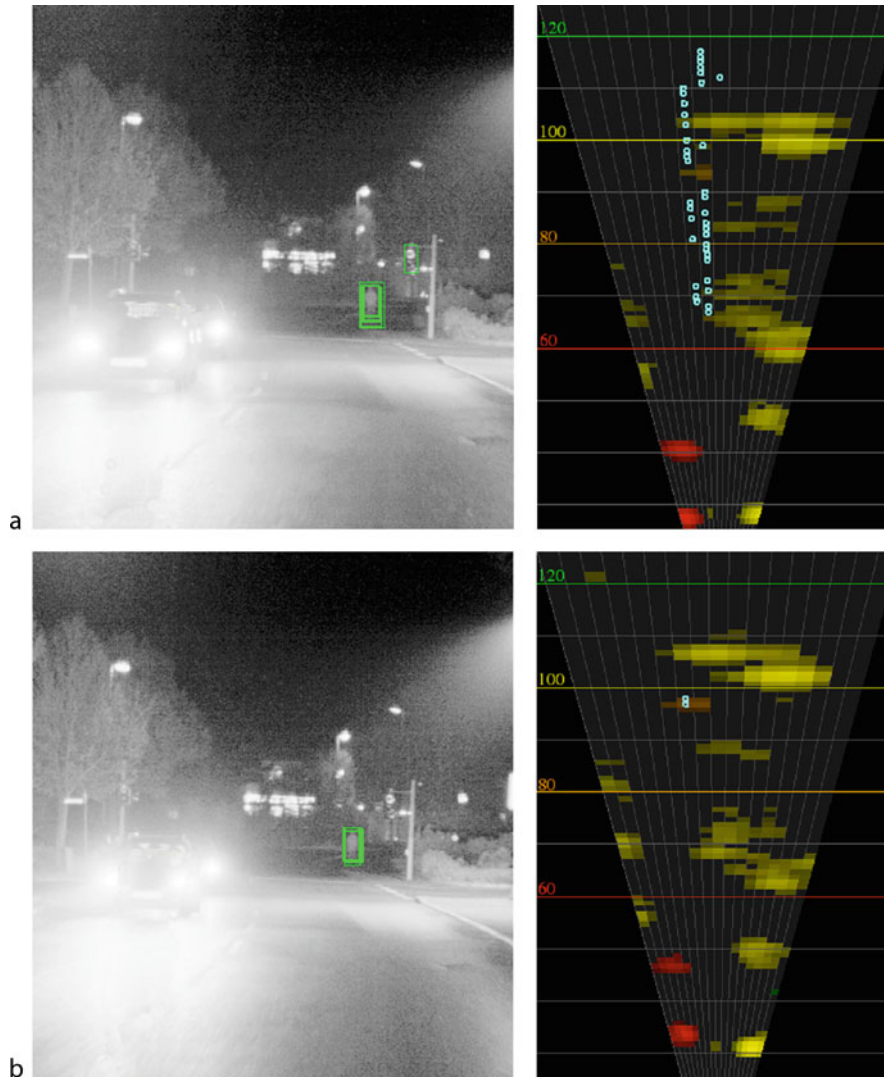
Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 18

Computational cost analysis in the early stages of the cascade up to layer 12: The image based reference detector and the fusion detector classify the same set of hypotheses. In stage one both detectors analyze about 2 million hypotheses using five weak learners (referenced as cost 100%). Due to the combination of radar features and image features, the fusion detector can reject more hypotheses than the image-based reference detector. Thus, the number of weak learner calculations can be reduced more efficiently. The fusion detector also shows similar results in highly structured environments, such as suburban areas with buildings and heavily changing light conditions



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 19

Performance comparison between the image based reference detector (solid) and the fusion detector (*dash-dot*) for the coverage 0.5 (blue, upper two graphs) and 0.6 (cyan, lower two graphs)

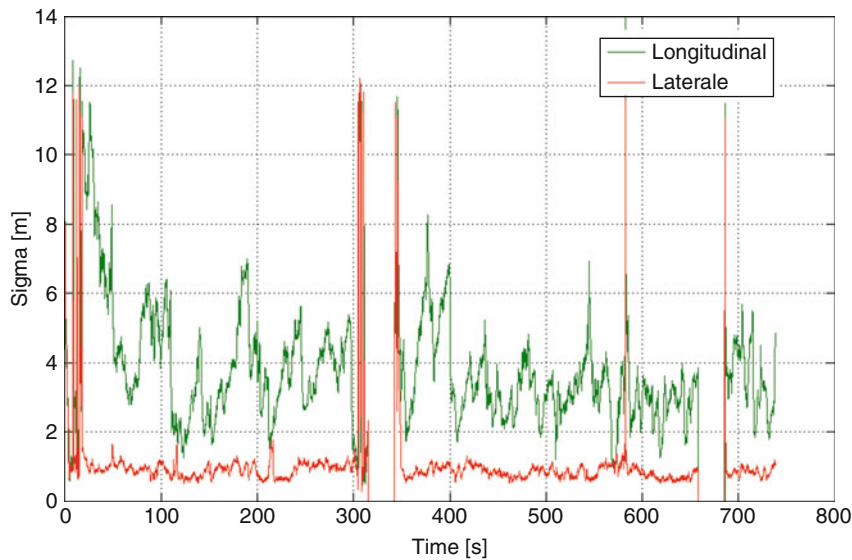


Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 20

Detection examples in a suburban environment (camera image with green detections left and radar data with cyan detections right): **(a)** Typical result of the image based reference detector. The uncertainty in spatial localization is very high using only image features for classification. Figure **(b)** shows the increased performance of the fusion detector incorporating features provided by the radar system too. The result is a precise localization of the pedestrians at 100 m ((a) Image-based reference detector: Typical detection of a pedestrian at 100 m with very imprecise spatial localization (70–120 m). (b) Fusion detector: Accurate detection of a pedestrian at 100 m)

The best, but unfortunately the most expensive, equipment is clearly a RTK GPS measurement unit. It is able to measure the vehicle position with a very high accuracy within a range of a few centimeters. For this examination, such a device was not available.

Nevertheless, the estimated position can also be analyzed considering the position uncertainty in a driving scenario. The position uncertainty calculation transforms each position hypothesis into the vehicle coordinate system of the expected position value of



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 21

Standard deviation of the estimated longitudinal and the lateral position for three driving scenarios

the distribution and determines the standard deviation for longitudinal and lateral position as well as for the orientation angle. Figure 21 shows the lateral and the longitudinal deviation for three scenarios. Whereas the lateral position deviation remains low and nearly constant, the longitudinal position deviation fluctuates more strongly. This behavior is associated with the current road course. A course with frequent curves provides more precise information for longitudinal localization than a straight road, which yields to ambiguous position hypotheses because the digital map fits anywhere along the road.

Moreover, the map-based position estimation can be compared to an approach using only the vehicle data that is the result of the egomotion estimation without resetting the position regularly. The better performance of the presented map matching algorithm against the egomotion approach is shown in Fig. 22.

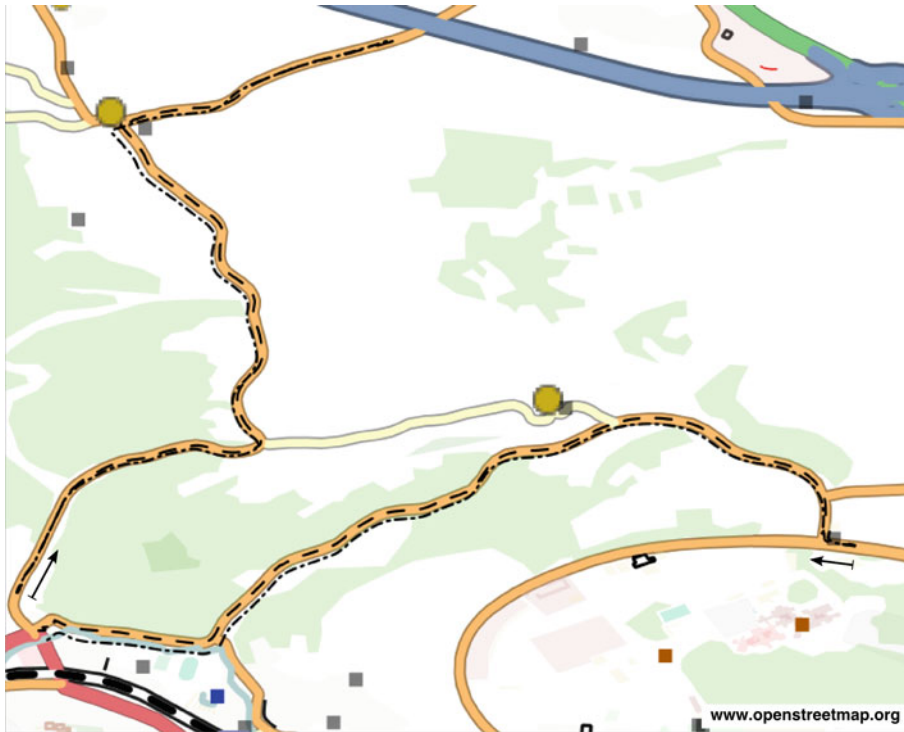
Note that the collection of investigated scenarios has been checked manually to diminish the impact of errors or outdated information from the digital map which cannot be corrected by the system itself. Correct information of the digital map is a necessary prerequisite condition of the realized road course estimation system.

Threat Analysis of Pedestrians The most interesting part is the evaluation of the warning performance answering the question: How many pedestrian detections have been classified correctly according to their relevance for the driver? This has been investigated on a ground truth of 1,945 labeled pedestrians in a range from 20 to 120 m with manually assigned discrete warning levels. Figure 23 shows the different warning levels. Manual labeling of the warning levels in the camera image is partially very complicated, mostly for detections in the far field of curved roads.

In order to reduce the resulting ground truth errors, the warning levels overlap between adjacent levels by 6%. This corresponds to a tolerance band of 0.4 m for a typical road width of 6.5 m. The evaluation algorithm classifies a detection considered to be correctly warned if at least one detection exists satisfying the following conditions.

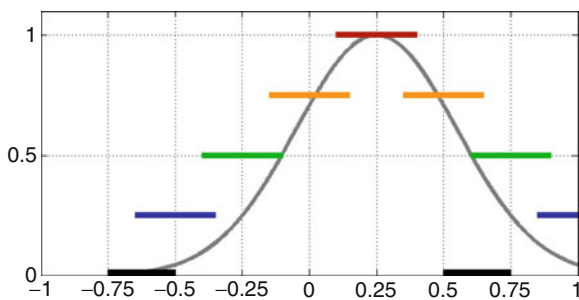
1. The calculated warning level $\eta(d^*)$ is in the specified range defined by the ground truth label.
2. The coverage between the ground truth label and the detection is at least 80%.

The presented pedestrian warning system achieves 84% correct warnings. The remaining 16% mostly refer to a position error of the presented road course estimation system.



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 22

Position trajectory with and without localization on the map: The course using only the vehicle data in an egomotion filter (*slash-dotted*) deviates clearly from the course of the map-based positioning algorithm (*dashed*)



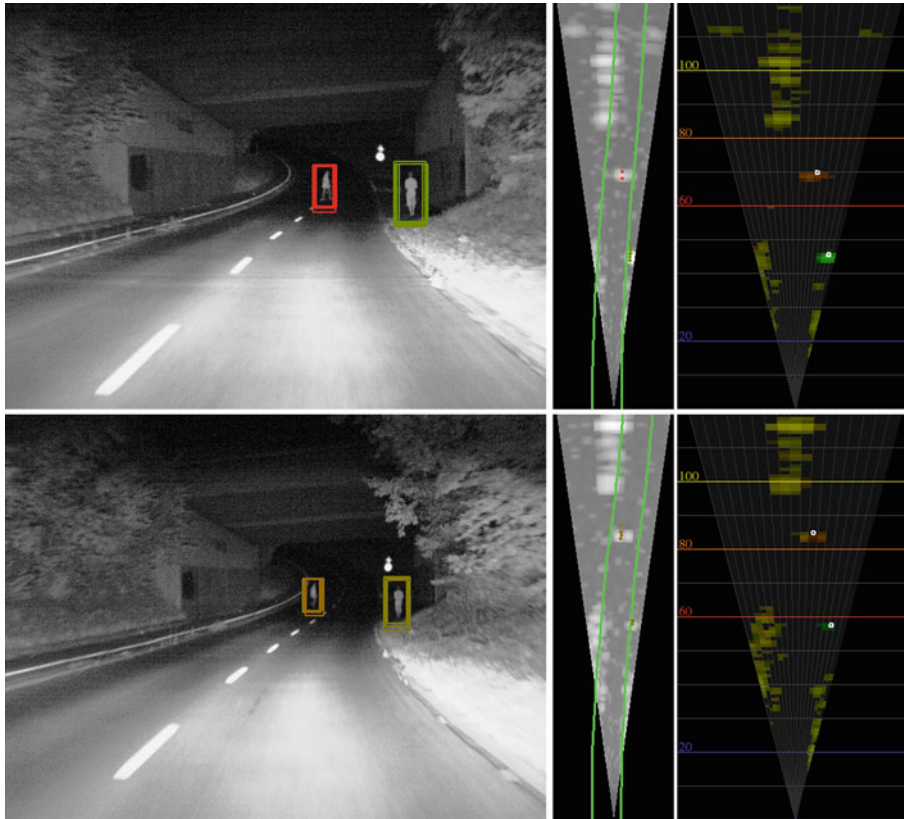
Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 23

The defined discrete warning levels

Conclusion

The described novel pedestrian warning system has been implemented in a test vehicle equipped with an NIR camera sensor, an imaging radar system, and a digital map.

In order to achieve a robust, reliable, and real-time capable detection system, the information provided by the camera sensor and the radar sensor are fused on feature level using a cascade of boosted classifiers. That is, both sensors define a set of features. For the classification task, the relevant feature are selected and parameterized using the supervised training algorithm AdaBoost in multiple stages of increased complexity (number of features). The feature set is composed of Haar-wavelet filters in the image plane and average object intensity as well as complex signal filters in the radar plane. The newly derived radar filters are characterized by their efficiency to separate between objects with a local velocity similarity and background. Experiments have shown that the resulting fusion detector dramatically outperforms a single sensor reference system. Unlike the image-based reference system, the presented multi-sensor fusion approach can reject hypotheses in earlier stages of the cascade and gives a precise spatial localization of the resulting object



Night Vision Pedestrian Warning in Intelligent Vehicles. Figure 24

Some results of the pedestrian warning system: the *left* camera image with highlighted pedestrians depending on their calculated threat. The next image corresponds to the interpolated Cartesian view of the radar raw data which is the input data to estimate the road course. The *right* image shows the radar raw data including the detected pedestrians

hypotheses. In this way, the fusion system also achieves a grid-based fusion with the digital map and the radar image.

In order to signal to the driver only the relevant objects, a new road course estimation concept has been implemented. The new concept maps the information provided by a digital map to the observations of the radar sensor, similar to the Monte Carlo localization approach in the field of autonomous robotics. Thus this component is more likely a global positioning approach which is apart from the digital map module, independent of inaccurate and error prone GPS measurements. Finally, the implemented threat analysis algorithm combines the pedestrian detection system and the road course estimation system. The evaluation results show a good warning performance. About 84%

of detected pedestrians triggered a correct warning. [Figure 24](#) gives an impression of the pedestrian warning system in a rural environment.

Future Directions

Future work will incorporate object classes other than pedestrians and more sensors to improve the environmental perception of vehicles, and will enhance the proposed approach.

The feature-level fusion for sensors of different physical principle is a possible solution to meet the requirements of safety application where high-level fusion is not sufficient. The described method is applicable not only for automotive applications, but also for example in the areas of monitoring, inspection, surveillance, and medicine.

Digital maps are currently the exclusive technology to provide information of the course of road for far distances over 100 m independent of weather conditions, which is necessary to realize a warning system. Today, neither a camera nor a radar sensor can provide this information. This shows that the future development of driver assistance systems should focus more on digital map technologies.

Fusion of local maps with digital maps will offer a lot of applications where a high-precision localization is necessary, and so far no low-cost solution is available.

Acknowledgments

This research was supported by PROPEDES (Predictive Pedestrian Protection at Night), BMBF FKZ 13N9750-13N9754, Germany.

Bibliography

Primary Literature

- Green M (2000) 'How long does it take to stop?' Methodological analysis of driver perception times. *Transp Hum Factors* 2(3):195–216
- Zhang Y, Antonsson E, Grote K (2006) A new threat assessment measure for collision avoidance systems. In: International conference on intelligent transportation systems (ITSC), 2006 IEEE, Toronto, pp 968–975.
- Beutnagel-Buchner U, Arnon M, Sparbert J, Moisel J, Rode M, Ritter W, Oberländer M, Löhlein O, Serfling M, Strobel M, Graf HG, Valipour H, Müller F, Schweiger R (2007) N I R W A R N Gesamtschlussbericht, BMBF, Forschungsbericht 01M3157A-D
- Franke U, Loose H, Knöppel C (2007) Lane recognition on country roads. In: Intelligent vehicles symposium (IV), 2007 IEEE, Istanbul, pp 99–104
- Jung CR, Kelber CR (2005) Lane following and lane departure using a linear-parabolic model. *Image Vis Comput* 23(13):1192–1202
- Apostoloff N, Zelinsky A (2003) Robust vision based lane tracking using multiple cues and particle filtering. In: Intelligent vehicles symposium (IV), 2003 IEEE, Columbus, pp 558–563
- Southall J, Taylor C (2001) Stochastic road shape estimation. In: International conference on computer vision (ICCV), 2001 IEEE, vol 1. Vancouver, pp 205–212
- Franke U (1992) Real-time 3D road modeling for autonomous vehicle guidance. In: Johansen P, Olsen S (eds) Theory and applications of image analysis: selected papers from the 7th Scandinavian conference on image analysis, Aalborg, Denmark, vol 2, Machine perception artificial intelligence. World Scientific Publishing Company, Singapore, pp 277–284
- Dickmanns ED, Zapp A (1986) A curvature-based scheme for improved road vehicle guidance by computer vision. In: Conference on mobile robots, 1986 SPIE, vol 727. Cambridge, pp 161–168
- Bertozzi M, Bombini L, Cerri P, Medici P, Antonello P, Miglietta M (2008) Obstacle detection and classification fusing radar and vision. In: Intelligent vehicles symposium (IV), 2008 IEEE, Eindhoven, Netherlands, pp 608–613
- Alessandretti G, Broggi A, Cerri P (2007) Vehicle and guard rail detection using radar and vision data fusion. *IEEE Trans Intell Transp Syst* 8(1):95–105
- Haselhoff A, Kummert A, Schneider G (2007) Radar-vision fusion with an application to car-following using an improved AdaBoost detection algorithm. In: Intelligent transportation systems conference (ITSC), 2007 IEEE, Seattle, pp 854–858
- Dehesa M, Allezard N, Furca L, Kaiser A, Kirchner A, Meinecke MM, Vinciguerra R (2005) SAVE-U: high and low level data fusion algorithms, Public Deliverable 19, 2005. Commission of the European Communities, Brussels
- Enzweiler M, Gavrila DM (2009) Monocular pedestrian detection: survey and experiments. *IEEE Trans Pattern Anal Mach Intell* 31(12):2179–2195
- Milch S, Behrens M (2001) Pedestrian detection with radar and computer vision. In: Symposium on progress in automobile lighting, Darmstadt, Germany, pp 657–664
- Schweiger R, Hamer H, Löhlein O (2007) Determining posterior probabilities on the basis of cascaded classifiers as used in pedestrian detection systems. In: Proceedings of IEEE intelligent vehicles symposium, Istanbul, Turkey, pp 1284–1289
- Kallenbach I, Schweiger R, Palm G, Löhlein O (2006) Multi-class object detection in vision systems using a hierarchy of cascaded classifiers. In: Proceedings of IEEE intelligent vehicles symposium, Tokyo, pp 383–387
- Miyamoto R, Sugano H, Nakamura Y (2007) Pedestrian recognition suitable for night vision systems. *Int J Comput Sci Netw Secur* 7:1–8
- Maehlich M, Oberländer M, Löhlein O, Gavrila D, Ritter W (2005) A multiple detector approach to low-resolution FIR pedestrian recognition. In: Proceedings of IEEE intelligent vehicles symposium, Las Vegas, pp 325–330
- Milch S, Behrens M (2001) Pedestrian detection with radar and computer vision. In: Proceedings of the symposium on progress in automobile lighting, Darmstadt, Germany, pp 657–664
- Lindl R, Walchshusl L (2006) Three-level early fusion for road use detection. *PREVENT Fusion Forum J* 1(1):19–24
- ProFusion (2004) Subproject of the PREVENT-project (Preventive and Active Safety Applications), founded by European Commission
- Dickmann J et al (2009) Sensorfusion as key technology for future driver assistance systems. In: Workshop Sensor- und Datenfusion – Architekturen und Algorithmen, Berlin, 20 Nov 2009
- Zhang Z (1999) Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the

international conference on computer vision, Kerkyra, Corfu, pp 666–673

25. Bouguet J (1999) Visual methods for three-dimensional modeling. Ph.D. dissertation, California Institute of Technology, Pasadena, CA
26. Tsai R (1987) A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE J Robot Automation* 3(4):323–344
27. Skolnik MI (1990) Radar handbook, 2nd edn. McGraw-Hill, New York
28. Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55:119–139
29. Viola P, Jones M (2002) Robust real-time object detection. *Int J Comput Vision* 57(2):137–154
30. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, Kauai, HI, vol 1, pp 511–518
31. Freund Y, Schapire RE (1996) Experiments with a new boosting algorithm. In: *Proceedings of the international conference on machine learning*, Bari, Italy, pp 148–156
32. Freund Y, Schapire RE (1999) A short introduction to boosting. *J Jpn Soc Artif Intell* 14(5):771–780
33. Schapire RE (2002) The boosting approach to machine learning: an overview. In: *MSRI workshop on nonlinear estimation and classification*, Berkeley, CA
34. Papageorgiou C, Oren M, Poggio T (1998) A general framework for object detection. In: *Proceedings of international conference on computer vision*, Bombay, India, pp 555–562
35. Oren M, Papageorgiou C, Sinha P, Osuna Z, Poggio T (1997) Pedestrian detection using wavelet templates. In: *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, San Juan, Puerto Rico, pp 193–199
36. Dellaert F, Fox D, Burgard W, Thrun S (1999) Monte Carlo localization for mobile robots. In: *International conference on robotics and automation (ICRA) 1999*, Detroit, vol 2. IEEE, Washington, DC, pp 1322–1328
37. Fox D, Burgard Z, Dellaert F, Thrun S (1999) Monte Carlo localization: efficient position estimation for mobile robots. In: *National conference on artificial intelligence (AAAI)*, Orlando, pp 343–349
38. Isard I, Blake A (1998) Condensation – conditional density propagation for visual tracking. *Int J Comput Vis (IJCV)* 29(1):5–28
39. Snyder JP (1987) Map projections – a working manual. US Government Printing Office, Washington, DC
40. Ress C, Balzer D, Bracht A, Durekovic S, Löwenau J (2008), ADASIS protocol for advanced in-vehicle applications. In: *World congress on intelligent transportation systems (ITS) 2008*, New York. IEEE, Washington, DC
41. Schubert R, Richter E, Wanielik G (2008) Comparison and evaluation of advanced motion models for vehicle tracking. In: *International conference on information fusion 2008*, Cologne, Germany. IEEE, Washington, DC, pp 1–6
42. Bierman GJ (1975) Measurement updating using the U-D factorization. In: *Conference on decision and control 1975*, Houston. IEEE, Washington, DC, pp 337–346

Books and Reviews

- Bar-Shalom Y, Li XR, Kirubarajan T (2001) Estimation with applications to tracking and navigation, 1st edn. Wiley, New York
- Blervaque V, Mezger K, Beuk L, Löwenau J (2006) ADAS horizon – How digital maps can contribute to road safety. In: Valldorf J, Gessner W (eds) *Advanced microsystems for automotive applications*. Springer, Berlin, pp 427–436
- Grabham D (2009) Street to screen: how sat-nav maps are made. *SatNav News* on www.techradar.com, August 2009
- Hager JW, Behensky JF, Drew BW (1989) The universal grids: universal transverse mercator (UTM) and universal polar stereographic (UPS). Technical manual DMATM 8358.2, Defense Mapping Agency (DMA), Fairfax, VA
- Misra P, Enge P (2006) *Global positioning system: signals, measurements and performance*, 2nd edn. Ganga-Jamuna Press, Lincoln
- Moran T (2008) Tapping mapping for an extra eye on the road. *Automotive News: Information Technology*, Detroit. July 2008

Non-flame Incineration

PHILIP H. TAYLOR

Environmental Engineering Group, University of
Dayton Research Institute, Dayton, OH, USA

Article Outline

- Glossary
- Definition of the Subject
- Introduction
- Thermal Stability–Based (UDRI) Incinerability Ranking
- Zone Theory of Formation of Products of Incomplete Combustion
- Formation of Chlorinated Polycyclic Aromatic Hydrocarbons
- Formation of Polychlorinated Dibenzo-P-Dioxins and Furans
- Future Directions
- Bibliography

Glossary

De novo synthesis of PCDD/F Synthesis of PCDD/F from native, nonextractable carbon in fly ash.

Molecular growth Molecular growth is defined as the increase in the molecular size and weight of products of incomplete combustion through a

series of radical–molecule reactions typically under oxygen-starved conditions.

Non-flame incineration Thermal destruction of hazardous compounds in the absence of a flame.

Precursor pathway of PCDD/F formation Formation of PCDD/F from the reaction of organic species derived from the combustion process, e.g., chlorophenols, on a transition metal surface.

Thermal stability index The thermal stability of an organic compound, defined by the temperature for a given level of conversion for a constant residence time of exposure, e.g., $T_{99}(2)(C)$ – the temperature for 99% conversion for a two-second gas-phase residence time.

Zone theory of PIC formation States that one can classify the types of reactions occurring within a given zone of a thermal destruction facility and consider their impact on PIC emissions in more detail.

Definition of the Subject

Non-flame incineration is an important and mature area of research pertaining to the thermal destruction of hazardous compounds. Research and field testing was initiated in the mid-1970s and has continued to this day. Non-flame incineration is defined as the “thermal” destruction of hazardous compounds in the absence of a flame. While flame conditions involve very high temperatures and very rapid reaction rates resulting in rapid conversion of organic compounds to thermodynamic end products (CO_2 , H_2O , HCl), non-flame incineration is typically defined by temperatures below 1,373 K in flowing air with mean, gas-phase residence times of 2.0 s. Non-flame incineration studies thus typically represent the gas-phase, rate-controlling conditions for pollutant formation in full-scale incinerators, where various failure modes in the flame zones result in the incomplete conversion of parent materials to thermodynamic end products. Non-flame incineration conditions are often similar to the second stage of a liquid injection or rotary kiln incineration system, which is often called an “afterburner.”

This review presents an update on non-flame incineration research. Topics covered include the continued development of a thermal stability–based ranking of principal organic hazardous constituents which has

been successfully used in full-scale trial burns and an update on research pertaining to the formation of products of incomplete combustion (PIC), with an emphasis on new findings reported over the past 20 years. The majority of the recent research in this latter area pertains to the formation and control of emissions of polychlorinated dibenzo-p-dioxins and furans (PCDD/F) from full-scale facilities. This research has been very successfully translated to full-scale thermal treatment facilities where overall PCDD/F emissions have decreased dramatically over the past 20 years.

Introduction

Controlled, high-temperature incineration, in spite of the associated high costs, is a proven organic waste reduction technology [1]. The current performance requirement states that principal organic hazardous constituents (POHC) designated in each waste must be destroyed and/or removed to an efficiency of 99.99%. The complexity of hazardous organic waste streams often makes correct POHC selection and demonstration of 99.99% destruction and removal efficiency (DRE) a nontrivial task.

To aid the process of POHC selection, the U.S. EPA has limited the number of hazardous organic compounds to those listed in Appendix VIII of 40 CFR Part 261.3 [2] and ranked these substances by their heat of combustion per unit mass ($\Delta H_c/g$) [3]. This scale is based on the premise that the lower the heat of combustion, the more difficult the compound is to incinerate. The heat of combustion approach has undergone considerable policy debate and received criticism on theoretical grounds from the scientific community [4–6].

The destruction efficiency (DE) of POHCs is largely controlled by the temperature, time, and fuel/air stoichiometry (excess air) in the high-temperature zones of incinerators. Calculations and experimental observations have shown that the emissions of undestroyed, residual POHCs are kinetically, not thermodynamically, controlled [5, 7]. Thus, determination of the exact time, temperature, and stoichiometry history of all the molecules in an incinerator is necessary to determine the absolute DE of a POHC. This type of information is not currently available.

However, sufficient information is available to estimate the relative DE of potential POHCs.

Simple conceptual and more complex models suggest that the gas-phase residence time and temperature in the post-flame or thermal zones of incinerators control the relative emissions of most POHCs [6, 8]. The basic theory or advanced hypothesis is that all molecules entering the flame zone of an incinerator are essentially destroyed, and only the minute fraction escaping the flame zone may be emitted from the facility. Various flame zone “failure modes” exist which may cause residual POHCs to be emitted. Once in the post-flame zone, gas-phase, non-flame, thermal decomposition kinetics control the rate of POHC destruction and PIC formation and destruction.

If all POHCs in a given waste stream in a given incinerator are volatilized at nearly the same rate, they will experience the same post-flame gas-phase residence time, temperature, and stoichiometry history. This means that the gas-phase thermal stability of the POHCs (as determined under a standardized set of conditions) may be used to predict their relative incinerability. The temperature for 99% destruction at 2.0 s gas-phase residence time, $[T_{99}(2)(C)]$, is one method of ranking the thermal stability of POHCs. Other residence times or temperatures may be used to develop this ranking. However, laboratory data indicate that although absolute POHC DEs are dependent upon time and temperature, the relative DEs are relatively insensitive to these parameters [9–11]. In contrast, reaction stoichiometry has been shown to be a major variable in determining relative thermal stability [10, 11].

Theoretical considerations suggest that oxygen-starved pathways through the incinerator are responsible for most POHC and PIC emissions. Even though a hypothetical facility is operating under nominally excess air conditions, poor mixing will result in oxygen-deficient pockets where the rate of POHC destruction is low and PIC formation is favored. Consequently, gas-phase thermal stability under sub-stoichiometric reaction conditions becomes an effective predictor of POHC relative incinerability. Dellinger et al. [12] compared the incinerability predictions of seven proposed POHC ranking methods with results of 10 pilot-scale or full-scale test burns. The ranking methods included heat of combustion,

autoignition temperature, ignition delay time, flame failure modes, theoretical flame mode kinetics, non-flame thermal stability of pure compounds under excess air conditions, non-flame thermal stability of mixtures under oxidative conditions, and the non-flame, thermal stability of mixtures under oxygen-starved conditions. Rank-order correlations of the prediction of the rankings with field results were poor except for non-flame thermal stability of mixtures under oxygen-starved conditions. Although the laboratory data base used to predict full-scale POHC DREs was limited, statistically significant correlations in seven of ten cases were observed using this ranking approach. The results of this comparison along with theoretical considerations suggest that this type of thermal stability ranking should be used to replace the previously recommended heat of combustion index.

Thermal Stability-Based (UDRI) Incinerability Ranking

Table 1 presents the updated thermal stability-based incinerability ranking of Appendix VIII organic compounds (with thermal stability indices) based on the gas-phase thermal stability under oxygen-starved conditions. The ranking was developed using experimental data generated at the University of Dayton Research Institute (UDRI), limited studies reported in the published literature, and theoretical considerations. The ranking (without thermal stability indices and class divisions) was initially published in 1990 [13]. Subsequent testing added many high-priority compounds including a number of polychlorinated dibenzo-p-dioxins and furans (PCDD/F). The ranking was subsequently validated with pilot and full-scale tests [14, 15].

In total, 86 compounds have been experimentally evaluated under a consistent set of oxygen-starved, low-chlorine experimental conditions. These compounds, where experimental data have been generated which is consistent with available literature data or theoretical consideration, are designated, experimentally evaluated, and are listed in boldface in Table 1. Compounds which have not been experimentally evaluated under these consistent test conditions, but for which experimental data exist within the published

Non-flame Incineration. Table 1 Updated UDRI thermal stability index

Principal organic hazardous constituent	T ₉₉ (2)(°C)	Rank
Class 1		
Hydrogen cyanide {hydrocyanic acid}^a	>1,500	1
Cyanogen {ethanedinitrile}	1,500	2
Benzene	1,150	3
Sulfur hexafluoride ^b	1,090	4
Naphthalene	1,070	5
Fluoranthene {benzo[j,k]fluorene}	1,062	6
Benzo[b]fluoranthene {2,3-benzofluoranthene}	1,060	7–11
Benanthracene (1,2-) {benz[a]anthracene}	1,060	7–11
Chrysene {1,2-benzphenanthrene}	1,060	7–11
Benzo[j]fluoranthene {7,8-benzofluoranthene}	1,060	7–11
Indeno(1,2,3-cd)pyrene {1,10-(1,2-phenylene)pyrene}	1,060	7–11
Benzo[a]pyrene {1,2-benzopyrene}	1,025	12
Dibenz[a,h]anthracene {1,2,5,6-dibenzanthracene}	1,020	13–16
Dibenzo[a,h]pyrene {1,2,5,6-dibenzopyrene}	1,020	13–16
Dibenzo[a,i]pyrene {1,2,7,8-dibenzopyrene}	1,020	13–16
Dibenzo[a,e]pyrene {1,2,4,5-dibenzopyrene}	1,020	13–16
<i>Chlorotrifluoromethane</i> ^{c, d}	1,010	17
Cyanogen chloride {chlorine cyanide}	1,000	18–19
Acetonitrile {ethanenitrile}	1,000	18–19
Chlorobenzene	990	20
Acrylonitrile {2-propenenitrile}	985	21
Dichlorobenzene {1,4-dichlorobenzene}	975	22–23
Chloronaphthalene (1-)	975	22–23
Cyanogen bromide {bromine cyanide}	970	24–25
Dichlorobenzene {1,2-dichlorobenzene}	970	24–25
Dichlorobenzene {1,3-dichlorobenzene}	960	26–27
Chlorodibenzo-P-dioxin (1-)	960	26–27
Trichlorobenzene (1,2,4-trichlorobenzene)	955	28–29
Tetrachlorobenzene (1,2,3,5-tetrachlorobenzene)^c	955	28–29
Trichlorobenzene (1,3,5-trichlorobenzene)^c	950	30–31
Chloromethane {methyl chloride}	950	30–31
Tetrachlorobenzene (1,2,4,5-tetrachlorobenzene)	940	32
Pentachlorobenzene	935	33–35
Hexachlorobenzene	935	33–35

Non-flame Incineration. Table 1 (Continued)

Principal organic hazardous constituent	T ₉₉ (2)(°C)	Rank
Bromomethane {methyl bromide}	935	33–35
Chlorodibenzo-P-dioxin (2-)	925	36
Dichlorodibenzo-P-dioxin (2,7-)	920	37
Class 2		
Toluene {methylbenzene}	895	38
Tetrachloroethene	890	39
Dde{1,1-dichloro-2,2-bis(4-chlorophenylethylene)}	875	40–41
Dichlorodibenzo-P-dioxin (2,3-)	875	40–41
Formic acid {methanoic acid}	870	42–43
Phosgene {carbonyl chloride}	870	42–43
Trichloroethene	865	44–45
Tetrachlorodibenzo-P-dioxin (2,3,7,8-)	865	44–45
Trichlorodibenzo-P-dioxin (1,2,4-)	860	46–48
Dichloroethene (1,1-)	860	46–48
Fluoroacetic acid	860	46–48
Dimethylbenz[a]anthracene (7,12-)	855	49
Aniline	850	50–53
Formaldehyde {methylene oxide}	850	50–53
Methyl chlorocarbonate {carbonochloridic acid, methyl ester}	850	50–53
Methyl isocyanate {methylcarbamide}	850	50–53
Tetrachlorodibenzo-P-dioxin (1,2,3,4-)	845	54
Dimethylphenethylamine (alpha, alpha-)	840	55–58
Malononitrile {propanedinitrile}	840	55–58
Naphthylamine (1-)	840	55–58
Naphthylamine (2-)	840	55–58
Dichloroethene (trans-1,2-)	835	59
Fluoroacetamide (2-)	830	60
Acrylamide {2-propenamide}	820	61–62
Methyl methacrylate {2-propenoic acid, 2-methyl-,methyl ester}	820	61–62
Dichloromethane {methylene chloride}	815	63–64
Methacrylonitrile {2-methyl-2-propenenitrile}	815	63–64
Chloroaniline {chlorobenzenamine}	810	65
Methylcholanthrene (3-)	805	66
Chloro-1,3-butadiene (2-) {chloroprene}	800	67–70
Diphenylamine {N-phenylbenzenamine}	800	67–70

Non-flame Incineration. Table 1 (Continued)

Principal organic hazardous constituent	T ₉₉ (2)(°C)	Rank
Pronamide {3,5-dichloro-N-[1,1-dimethyl-2-propynyl] benzamide}	800	67–70
Acetylaminofluorene (2-) {acetamide,N-[9 H-fluoren-2-YL]-}	800	67–70
Class 3		
Aminobiphenyl (4-) {[1,1'-biphenyl]-4-amine}	796	71
Dichlorobenzidine (3,3'-)	795	72–73
Chlorophenol (2-)	795	72–73
Benzidine {[1,1'-biphenyl]-4,4'-diamine}	794	74
Dimethylbenzidine (3,3'-)	793	75
Phenylenediamine (1,3-) {benzenediamine}	792	76
Phenylenediamine (1,2-) {benzenediamine}	791	77
Phenylenediamine (1,4) {benzenediamine}	790	78
n-Propylamine {1-propanamine}	789	79
Pyridine	785	80–81
Chlorophenol (3-)	785	80–81
Picoline (2-) {pyridine,2-methyl-}	780	82–85
Dichloropropene (1,1-)	780	82–85
Thioacetamide {ethanethioamide}	780	82–85
Trichloro(1,2,2-)trifluoroethane(1,1,2-) {freon 113}^c	780	82–85
Phenol {hydroxybenzene}	775	86–90
Benz[c]acridine {3,4-benzacridine}	775	86–90
Dichlorodifluoromethane {freon 12}	775	86–90
Acetophenone {ethanone,1-phenyl-}	775	86–90
Trichlorofluoromethane {freon 11}	775	86–90
Ethyl cyanide {propionitrile}	770	91–93
Benzoquinone {1,4-cyclohexadienedione}	770	91–93
Vinyl chloride (chloroethene)	770	91–93
Dibenz[a,h]acridine {1,2,5,6-dibenzacridine}	765	94–99
Dibenz[a,j]acridine {1,2,7,8-dibenzacridine}	765	94–99
Hexachlorobutadiene (trans-1,3)	765	94–99
Naphthoquinone (1,4-) {1,4-naphthalenedione}	765	94–99
Dimethyl phthalate	765	94–99
Acetyl chloride {ethanoyl chloride}	765	94–99
Acetonylbenzyl-4-hydroxycoumarin (3- α -) {warfarin}	760	100–101
Maleic anhydride {2,5-furandione}	760	100–101
Chlorophenol (4-)	755	102–103
Dichloro-2-butene (1,2-)	755	102–103

Non-flame Incineration. Table 1 (Continued)

Principal organic hazardous constituent	T ₉₉ (2)(°C)	Rank
Dichloropropene (2,3-)	754	104
Dichloropropene (trans-1,2-)	750	105–113
Dibenzo[c,g]carbazole (7 h-) {3,4,5,6-dibenzcarbazole}	750	105–113
Toluenediamine (2,6-) {diaminotoluene}	750	105–113
Toluenediamine (2,5-) {diaminotoluene}	750	105–113
Toluenediamine (2,4-) {diaminotoluene}	750	105–113
Toluenediamine (2,4-) {diaminotoluene}	750	105–113
Toluenediamine (3,5-) {diaminotoluene}	750	105–113
Toluenediamine (3,4-) {diaminotoluene}	750	105–113
Resorcinol {1,3-benzenediol}	750	105–113
Cresol (1,3-) {methylphenol}	749	114
Cresol (1,2-) {methylphenol}	746	115
Cresol (1,4-) {methylphenol}	745	116
Dichlorophenol (2,6-)	740	117
Methyl ethyl ketone {2-butanone}	730	118–119
Diethylstilbesterol	730	118–119
Benzenethiol {thiophenol}	725	120
Isobutyl alcohol {2-methyl-1-propanol}	715	121
Crotonaldehyde {2-butenal}	710	122–123
Dichlorophenol (2,4-)	710	122–123
Chlorocresol {4-chloro-3-methylphenol}	709	124
Methylactonitrile (2-) {propanenitrile,2-hydroxy-2-methyl}	705	125–126
Dihydroxy- α -[methylamino]methyl benzyl alcohol (3,4-) {adrenaline}	705	125–126
Dimethylphenol (2,4-)	700	127
Class 4		
Chloropropene (3-) {allyl chloride}	695	128
Tetrachloroethane (1,1,2,2-)	690	129–130
Dichloropropene (3,3-)	690	129–130
<i>Hydrazine {diamine}</i>	685	131–133
Benzyl chloride {chloromethylbenzene}	685	131–133
Dibromomethane {methylene bromide}	685	131–133
Dichloroethane (1,2-)	680	134
Mustard gas {bis[2-chloroethyl]-sulfide}	679	135–137
Nitrogen mustard	679	135–137
N,N-bis(2-chloroethyl)2-naphthylamine {chlornaphazine}	679	135–137
Acrolein {2-propenal}	670	138–141

Non-flame Incineration. Table 1 (Continued)

Principal organic hazardous constituent	T ₉₉ (2)(°C)	Rank
Bromoacetone {1-bromo-2-propanone}	670	138–141
<i>Chloroethane (ethyl chloride) [4]</i>	670	138–141
Hexachlorophene {2,2'-methylenebis[3,4,6-trichlorophenol]}	670	138–141
Allyl alcohol {2-propen-1-ol}	665	142
Dioxane (1,4-) {1,4-diethylene oxide}	660	143
Chlorambucil	659	144
Nitrobenzene	655	145–146
Chloropropionitrile (3-) {3-chloropropanenitrile}	655	145–146
Dichloro-2-propanol (1,1-)	654	147–148
Ddd {dichlorodiphenyldichloroethane}	654	147–148
Dichloro-2-propanol (1,3-)	652	149
Phthalic anhydride {1,2-benzenedicarboxylic acid anhydride}	650	150–154
Methyl parathion	650	150–154
Nitrophenol (4-)	650	150–154
Trichlorophenol (2,4,6-)	650	150–154
Trichlorophenol (2,4,5-)	650	150–154
Tetrachloromethane {carbon tetrachloride}	645	155–157
Tetrachlorophenol (2,3,4,6-)	645	155–157
Chlorodifluoromethane {freon 22}^c	645	155–157
Hexachlorocyclohexane {lindane}	640	158–163
Dichlorofluoromethane {freon 21}^c	640	158–163
Dinitrobenzene (1,3-)	640	158–163
Nitroaniline {4-nitrobenzenamine}	640	158–163
<i>Pentachlorophenol</i>	640	158–163
Pentachloroethane	640	158–163
Dinitrobenzene (1,4-)	635	164–166
Dinitrobenzene (1,2-)	635	164–166
Trichloroethane (1,1,2-)	635	164–166
Isodrin	632	167–169
Dieldrin	632	167–169
Aldrin	632	167–169
<i>Dichloropropane (1,3-)</i>	630	170–174
Dibromoethane (1,2-) {ethylene dibromide}	630	170–174
Nitrotoluidine (5-) {benzenamine,2-methyl-5-nitro-}	630	170–174
Chloroacetaldehyde	630	170–174
Benzal chloride {alpha,alpha-dichlorotoluene}	630	170–174

Non-flame Incineration. Table 1 (Continued)

Principal organic hazardous constituent	T ₉₉ (2)(°C)	Rank
Trichloromethane {chloroform}	625	175–183
Dichloropropene (trans-1,3-)	625	175–183
Trichloropropane (1,2,3-)	625	175–183
Dinitrotoluene (2,4-)	625	175–183
Dinitrotoluene (2,6-)	625	175–183
Hexachlorocyclopentadiene	625	175–183
Dichloro-1-propanol (2,3-)	625	175–183
Ethylene oxide {oxirane}	625	175–183
Dichloropropene (cis-1,3-)	625	175–183
Dimethylcarbamoylchloride	620	184–187
Glycidyaldehyde {1-propanol-2,3-epoxy}	620	184–187
Ddt {dichlorodiphenyltrichloroethane}	620	184–187
Dichloropropane (1,2-) {propylene dichloride}	620	184–187
Auramine	618	188–189
Heptachlor	618	188–189
Dichloropropane (1,1-)	615	190–194
Chloro-2,3-epoxypropane (1-) {oxirane,2-chloromethyl-}	615	190–194
Dinitrophenol (2,4-)	615	190–194
bis(2-Chloroethyl)ether	615	190–194
Trinitrobenzene {1,3,5-trinitrobenzene}	615	190–194
Butyl-4,6-dinitrophenol (2-sec-) {DNBP}	613	195–196
Cyclohexyl-4,6-dinitrophenol (2-)	613	195–196
Dichloroethane (1,1-) {ethylidene dichloride}	610	197–201
bis(2-Chloroethoxy)methane	610	197–201
Chloral {trichloroacetaldehyde}	610	197–201
Trichloromethanethiol	610	197–201
Dinitroresol (4,6-) {phenol,2,4-dinitro-6-methyl-}	610	197–201
Heptachlor epoxide	609	202
Diepoxybutane (1,2,3,4-) {2,2'-bioxirane}	604	203
Benzotrichloride {trichloromethylbenzene}	600	204–205
Methapyrilene	600	204–205
Class 5		
Phenacetin {N-[4-ethoxyphenyl]acetamide}	593	206–207
Methyl hydrazine	593	206–207
Aflatoxins	590	208
Hexachloroethane	585	209–210

Non-flame Incineration. Table 1 (Continued)

Principal organic hazardous constituent	T ₉₉ (2)(°C)	Rank
Bromoform {tribromomethane}	585	209–210
Chlorobenzilate	584	211–214
Ethyl carbamate {urethan} {carbamic acid, ethyl ester}	584	211–214
Ethyl methacrylate {2-propenoic acid, 2-methyl-, ethyl ester}	584	211–214
Lasiocarpine	584	211–214
Amitrole {1 h-1,2,4-triazol-3-amine}	580	215–216
Muscimol {5-aminomethyl-3-isoazotol}	580	215–216
Iodomethane {methyl iodide}	567	217
Dichlorophenoxyacetic acid (2,4-) {2,4-d}	565	218–220
Chloroethylvinylether (2-) {ethene, [2-chloroethoxy]-}	565	218–220
Methylene bis(2-chloroaniline) (4,4-)	565	218–220
Dibromo-3-chloropropane (1,2-)	561	221
Propyn-1-OL (2-) {propargyl alcohol}	560	222–223
Tetrachloroethane (1,1,1,2-)	560	222–223
<i>Dimethylhydrazine (1,1-)</i>	555	224–225
N,N-diethylhydrazine {1,2-diethylhydrazine}	555	224–225
Chloromethylmethyl ether {chloromethoxymethane}	550	226–228
Dimethyl-1-methylthio-2-butanone,O-[(methylamino)-carbonyl]	550	226–228
Oxime (3,3-) {thiofanox}		
Dimethylhydrazine (1,2-)	550	226–228
Trichloroethane (1,1,1-) {methyl chloroform}	545	229
Chlordane (alpha and gamma isomers)	533	230
bis(Chloromethyl)ether {methane-OXYbis[2-chloro-]}	530	231–233
Parathion	530	231–233
<i>Dichloropropane (2,2-)</i>	530	231–233
Maleic hydrazide {1,2-dihydro-3,6-pyridazinedione}	525	234
Bromophenyl Phenyl Ether (4-) {benzene,1-bromo-4-phenoxy-}	523	235
bis(2-Chloroisopropyl)ether	520	236–237
Dihydrosafrole {1,2-methylenedioxy-4-propylbenzene}	520	236–237
Methyl methanesulfonate {methanesulfonic acid, methyl ester}	515	238
Propane sulfone (1,3-) {1,2-oxathiolane,2,2-dioxide}	513	239
Saccharin {1,2-benzisothiazolin-3-one,1,1-dioxide}	511	240
Methyl-2-methylthio-propionaldehyde-O-(methylcarbonyl)oxime (2-)	510	241–242
Methyomyl	510	241–242
Hexachloropropene	505	243
Pentachloronitrobenzene {PCNB}	500	244–248

Non-flame Incineration. Table 1 (Continued)

Principal organic hazardous constituent	T ₉₉ (2)(°C)	Rank
Diallate {s-(2,3-dichloroallyl)diisopropyl thiocarbamate}	500	244–248
Ethyleneimine {aziridine}	500	244–248
Aramite	500	244–248
Dimethoate	500	244–248
Trichlorophenoxyacetic acid (2,4,5-) {2,4,5-t}	495	249–250
Trichlorophenoxypropionic acid (2,4,5-) {2,4,5-TP} {SILVEX}	495	249–250
tris(2,3-Dibromopropyl)phosphate	492	251
Methylaziridine (2-) {1,2-propylenimine}	490	252–253
Methoxychlor	490	252–253
Brucine {strychnidin-10-one,2,3-dimethoxy-}	470	254–255
Kepone	470	254–255
Isosafrole {1,2-methylenedioxy-4-allylbenzene}	460	256–258
Safrole {1,2-methylene-4-allylbenzene}	460	256–258
tris(1-Azridinyl) phosphine sulfide	460	256–258
Dimethoxybenzidine (3,3'-)	453	259
Diphenylhydrazine (1,2-)	450	260
O,O-diethylphosphoric acid, O-p-nitrophenyl ester	425	261
n-Butylbenzyl phthalate	415	262
O,O-diethyl-O-2-pyrazinyl phosphorothioate	410	263
Dimethylaminoazobenzene	405	264
Diethyl phthalate	400	265–266
Class 6		
O,O-diethyl-S-methyl ester of phosphoric acid	395	265–266
O,O-diethyl S-[(ethylthio)methyl]ester of phosphorodithioic acid	395	267–268
Citrus red No. 2 {2-naphthol,1-[(2,5-dimethoxyphenyl)azo]}	395	267–268
Trypan blue	392	269
Ethyl methanesulfonate {methanesulfonic acid, ethyl ester}	390	270–274
Disulfoton	390	270–274
Diisopropylfluorophosphate {DFP}	390	270–274
O,O,O-Triethyl phosphorothioate	390	270–274
Di-n-butyl phthalate	390	270–274
Paraldehyde {2,4,6-trimethyl-1,3,5-trioxane}	380	275
Di-n-octyl phthalate	380	276
Octamethylpyrophosphoramidate {Octamethyldiphosphoramidate}	374	277
bis(2-Ethylhexyl)phthalate	370	278–279
Methylthiouracil	370	278–279

Non-flame Incineration. Table 1 (Continued)

Principal organic hazardous constituent	T ₉₉ (2)(°C)	Rank
Propylthiouracil	360	280
Strychnine {strychnidin-10-one}	320	281
Cyclophosphamide	300	282–285
Nicotine {(s)-3-[1-methyl-2-pyrrolidinyl]pyridine}	300	282–285
Reserpine	300	282–285
Toluidine hydrochloride {2-methyl-benzenamine hydrochloride}	300	282–285
Class 7		
Tolylene diisocyanate {1,3-diisocyanatomethylbenzene}	290	286
Endrin	285	287
Butanone peroxide (2-) {methyl ethyl ketone, peroxide}	260	288
Tetraethylpyrophosphate	255	289
Nitroglycerine [{trinitrate-1,2,3-propanetriol}]	255	290
Tetraethyldithiopyrophosphate	250	291–292
Ethylenebisdithiocarbamic acid	250	291–292
Tetranitromethane	245	293
Uracil mustard {5-[bis(2-chloroethyl)amino]uracil}	246	294
Acetyl-2-thiourea (1-) {acetamide,N-[aminothioxomethyl]-}	240	295–299
Chlorophenyl thiourea (1-) {thiourea,[2-chlorophenyl]-}	240	295–299
N-phenylthiourea	240	295–299
Naphthyl-2-thiourea (1-) {thiourea,1-naphthalenyl-}	240	295–299
Thiourea {thiocarbamide}	240	295–299
Daunomycin	230	300–301
Ethylene thiourea {2-imidazolidinethione}	230	300–301
Thiosemicarbazide {hydrazinecarbothioamide}	225	302–303
Melphalan {alanine,3-[p-bis(2-chloroethyl)amino] phenyl,-L-}	225	302–303
Dithiobiuret (2,4-) {thioimidodicarbonic diamide}	220	304–305
Thiuram {bis[dimethylthiocarbamoyl]disulfide}	220	304–305
AzaserinE {L-serine,diazoacetate [ester]}	200	306
Hexaethyl tetraphosphate	190	307
Nitrogen mustard N-oxide	170	308–310
Nitroquinoline-1-oxide (4-)	170	308–310
Cycasin {beta-D-glucopyranoside, [methyl-ONN-azoxy]methyl-}	170	308–310
Streptozotocin	155	311
N-methyl-N'-nitro-N-nitrosoguanidine	130	312–327
N-nitroso-di-ethanolamine {[2,2'-nitrosoimino]bisethanol}	130	312–327
N-nitroso-di-N-butylamine {N-butyl-N-nitroso-1-butanamine}	130	312–327

Non-flame Incineration. Table 1 (Continued)

Principal organic hazardous constituent	T ₉₉ (2)(°C)	Rank
N-nitroso-N-ethylurea {N-ethyl-N-nitrosocarbamide}	130	312–327
N-nitroso-N-methylurea {N-methyl-N-nitrosocarbamide}	130	312–327
N-nitroso-N-methylurethane	130	312–327
N-nitrosodiethylamine {N-ethyl-N-nitrosoethanamine}	130	312–327
N-nitrosodimethylamine {dimethylnitrosamine}	130	312–327
N-nitrosomethylethylamine {N-methyl-N-nitrosoethanamine}	130	312–327
N-nitrosomethylvinylamine {N-methyl-N-nitrosoethenamine}	130	312–327
N-nitrosomorpholine	130	312–327
N-nitrosornicotine	130	312–327
N-nitrosopiperidine {hexahydro-N-nitrosopyridine}	130	312–327
N-nitrososarcosine	130	312–327
Nitrosopyrrolidine {N-nitrosotetrahydropyrrole}	130	312–327
Di-n-propylnitrosamine {N-nitroso-di-n-propylamine}	130	312–327
Oxabicyclo[2.2.1]heptane-2,3-dicarboxylic acid (7-) {endothal}	120	328
endosulfan	100	329

^aBoldface print indicates compound thermal stability is fully evaluated; ranking based on UDRI experimental mixture data coupled with reaction kinetic theory

^bNon-appendix VIII compound

^cNOS listing; ranking is presented based on either UDRI or literature experimental data coupled with reaction kinetic theory

^dItalicized print indicates compound thermal stability is fully evaluated based on UDRI or literature data coupled with reaction kinetic theory

literature, are listed in italics. The remaining compounds were ranked based on limited experimental data on the actual compound, experimental data on similar compounds, and/or reaction kinetic theory. For this latter group, which includes the majority of the ranking, sufficient data were not available to designate the compounds as experimentally evaluated. Thus, their ranking is considered less accurate than the experimentally evaluated, bold-faced POHCs.

The ranking was devised by assigning T₉₉(2)(C) values to each compound. For some compounds, initially identical values of T₉₉(2)(C) were assigned resulting in identical rankings. For example, there was no clear-cut basis for predicting the relative stability of diphenylamine, 1,1-dichloroethene, and fluoroacetic acid. Hence, the three compounds received a “tie” ranking as the 41st most stable compound. In other instances, the compounds that were initially assigned identical, T₉₉(2)(C) values belonged to a homologous

series. In this case, the mechanisms of destruction were similar and relative reaction rates could be assigned. This was generally the case for polycyclic aromatic hydrocarbons (PAHs); consequently, unique rankings could be assigned for most of these compounds. In general, it was not possible to further distinguish the initial ranking when the compounds in question decomposed by different mechanisms. The uncertainty in predicting T₉₉(2)(C) values, the accuracy of field test data, and the variation in the actual pollutant emissions from facility to facility does not justify a more precise 1 to 1 ranking of the Appendix VIII organic compounds.

The ranking includes all individually listed organic compounds included on the Appendix VIII list. Appendix VIII also lists many compounds by class, i.e., not otherwise specified (NOS). These classes of compounds are generally not ranked. However, there are many instances where compounds are listed individually and by class. When this occurs, the ranking

includes the individually listed compounds. Exceptions to this rule were made for 1,2,2-trichloro-1,1,2-trifluoroethane, dichlorofluoromethane, chlorodifluoromethane, ethyl chloride, 1,3,5-trichlorobenzene, and 1,2,3,5-tetrachlorobenzene. Experimental data were available for these compounds, and they have been of interest as frequently selected POHCs or possible incineration surrogates. Sulfur hexafluoride has also been included in the ranking although it is not listed on Appendix VIII, due to interest in its utility as an incineration surrogate [16].

Application of the Updated UDRI Incinerability Ranking

The ranking in Table 1 is also grouped by horizontal lines which separate compounds into nine thermal stability classes, i.e., very stable, moderate stable, fragile, etc. This grouping is generally based on natural divisions in the predicted stability or reaction kinetic theory which suggests significantly different mechanistic pathways for the non-flame thermal decomposition of compounds within different classes.

One may use the thermal stability ranking in the same manner as the previous heat of combustion index, i.e., selection of a group of POHCs for trial burn testing and only allow POHCs to be burned which are lower in the ranking than the most fragile POHC for which 99.99% DRE can be demonstrated. An alternate approach would be to only allow trial burn testing of experimentally evaluated POHCs, but allow incineration of all POHCs lower in the ranking than the least stable POHC for which 99.99% DRE can be demonstrated.

Another alternate approach involves the use of the incinerability class divisions. In this approach, one can demonstrate 99.99% DRE of a group of experimentally evaluated POHCs within an incinerability class and receive approval to burn all other POHCs within and below that class. For example, if 99.99% DRE can be demonstrated for chloromethane, acetonitrile, and chlorobenzene (all class 1 compounds), the facility would be allowed to burn any Appendix VIII compound. As a second example, if 99.99% DRE is demonstrated for 1,3-hexachlorobutadiene, methyl ethyl ketone, and cresol (all class 3 compounds), the facility would only be allowed to burn compounds ranked in class 3 or lower.

It should be noted that reformation of a POHC from its own decomposition products or formation of the POHC as a PIC from another waste feed component is generally not included as a factor in this ranking. Significantly, laboratory studies have shown that both POHC reformation and PIC formation can affect observed DEs [11, 17]. In some instances, laboratory determination of $T_{99(2)}(C)$ is unavoidably affected by POHC reformation (e.g., formation of chloromethane from chlorine atom attack on methane). When it was judged that POHC reformation was inevitable from almost any waste, reformation was included in the thermal stability ranking assignment.

The extensive testing conducted by UDRI has shown that the ranking of most of the compounds is relatively insensitive to composition of the waste mixture. However, a few notable exceptions have been observed in initial tests. This includes aniline derivatives and PAHs in general. Somewhat lesser variability has been observed for benzene, toluene, and pyridine, which were judged to be sufficiently constant to be included as experimentally evaluated compounds. Consequently, it is suggested that the waste composition be carefully considered before these compounds are selected as POHCs. Benzene, toluene, and PAHs are expected to be particularly prone to formation as PICs in full-scale trial burn incineration testing, and observed DREs may be significantly affected. Effects of PIC formation may be minimized by feeding the POHC at a sufficiently high concentration.

Finally, it should be noted that no rankings are given for polychlorinated biphenyls (PCBs) which are listed as NOS in Appendix VIII. An estimate of the stability of individual PCB congeners may be made by assuming that the congener is slightly less stable than the chlorinated benzene corresponding to the least chlorinated ring in the PCB congener of interest. This means, for example, that dichlorobenzene would be a suitable incineration surrogate for 2,2',3,3',4-pentachlorobiphenyl. Similarly, chlorobenzene would be a suitable incineration surrogate for 3,3',4,5-tetrachlorobiphenyl, and pentachlorobenzene a suitable surrogate for decachlorobiphenyl.

Metal-containing compounds listed in Appendix VIII were not included in the ranking. In many cases,

the metal-containing compound is not converted in the high-temperature process and thus does not fit the description of DRE in the regulatory context.

Zone Theory of Formation of Products of Incomplete Combustion

Because of the wide varieties of physical and chemical compositions of hazardous waste, a universal treatment technique must have a very general mechanism of interaction with the waste. While chemical and photolytic treatment methods may be suitable for specific compounds or waste streams which are susceptible to a particular mechanism of chemical attack or absorb incident radiation, they are far from universal. Because all molecules absorb infrared radiation (viz. thermal energy), they are in principle, all susceptible to thermal treatment. It is largely for this reason that incineration has received such broad application. However, each type of molecule decomposes by a different mechanism, and the resulting fragments can undergo further reaction to form a wide variety of products. It is largely because of the potential complexity and toxicity of this “soup” of thermal by-products that incineration has received such intense scientific and regulatory scrutiny. In contrast to metals, which are not destroyed and undergo only a limited number of chemical reactions, it is feared that the situation concerning organic by-products may be too complex to ever solve, and one may unknowingly pollute the atmosphere with these “toxins.”

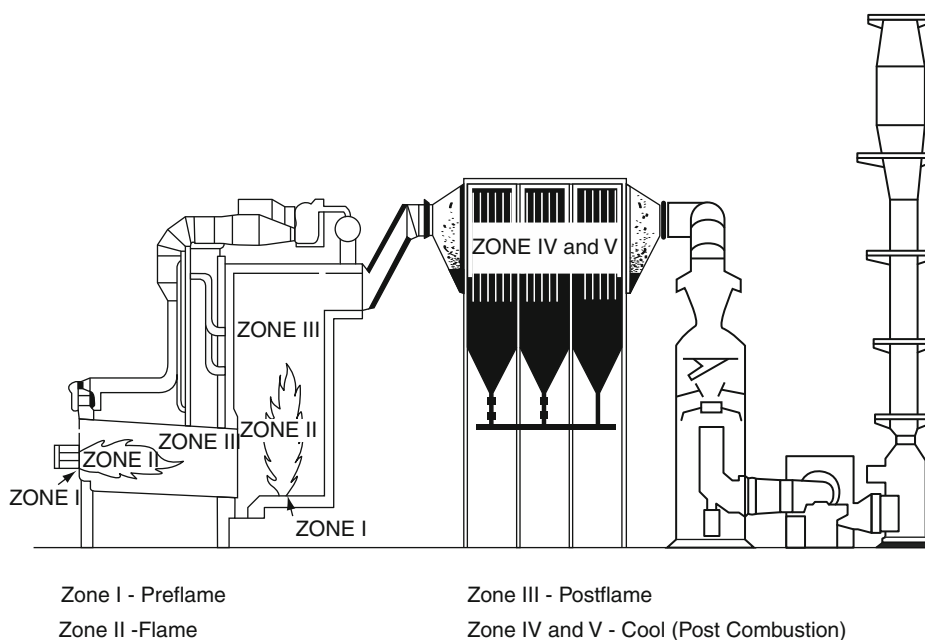
However, chemists have long known that chemicals react by a limited number of elementary mechanisms. Although the situation is still complex due to the number of compounds contained in a typical waste feed, the number of even trace by-products that can be formed is limited by these mechanisms. Thus, if sufficient effort is placed into identifying and classifying these reactions, we can predict most of the potential by-products. Using this same knowledge of chemistry, we can do an even better job of identifying those compounds that are found in the highest yields, are the most thermally stable, and thus most significantly challenge the performance of a given incineration system. Such an approach is highly beneficial because many toxic by-products may escape undetected unless they are targeted for highly sensitive chemical analysis.

There are many ways to categorize reactions that lead to the emission of organic combustion by-products. From an incineration viewpoint, it is most useful to classify reaction pathways in a manner that they can be associated with regions of the incinerator where they are the most significant. Figure 1 depicts a typical rotary kiln/afterburner-type unit with four reaction zones being designated.

Zone 1, the pre-flame zone, is characterized by a wide range of temperatures (near ambient to approaching combustion temperatures), residence times on the order of 0.1 s, and low-excess-air conditions. Because this zone occurs at the front end of the device, it is not considered to be directly responsible for a very large fraction of pollutant emissions. Instead, it can create new reactants, by several low energy pathways that react further in the downstream zones.

Zone 2, the flame zone, is characterized by temperatures of 1,273–2,073 K at which essentially every organic compound will undergo complete conversion to its most thermodynamically stable end products, e.g., CO₂, H₂O, HCl, NO, etc. Numerous studies of premixed and diffusion flames have shown this to be the case, with any observed emissions of organic species being due to flow-paths that pass through the periphery of the flame, follow eddies of poor fuel/air mixing, or circumvent the flame altogether [8, 18, 19]. These flow-paths represent destruction “failure modes” of the flame and generate pockets that are more properly described as high-temperature thermal zones, viz. Zone 3. Although the flame zone probably does not generate significant toxic organic pollutants, it does generate large quantities of some inorganic pollutants, e.g., NO and HCl, as well as other lower concentration species, e.g., Cl₂ and NO₂, that may be very important reactants in subsequent zones.

Zone 3, the high-temperature, non-flame thermal zone, is a chemistry-rich zone where many types of high-energy unimolecular and bimolecular radical-molecule reactions occur. It is characterized by temperatures from 873 to 1,373 K, residence times of a few seconds, and both oxygen-rich as well as oxygen-depleted regions. Experimental as well as modeling studies indicate that the majority of the pollutant formation in this zone occurs in oxygen-depleted pockets of poor waste-air mixing [8, 18–20]. It is within



Non-flame Incineration. Figure 1

Chemical reaction zones in a hazardous waste incinerator. *Zone I* – Pre-flame Zone, *Zone II* – Flame Zone, *III* – Post-flame Zone, *Zone IV & V* – Post-combustion (cool) zones

this zone that most of the polycyclic aromatic hydrocarbons (PAHs), heteroatom-containing polynuclear aromatic hydrocarbons (PNAs), and some higher molecular weight chlorinated hydrocarbons (CHCs) are formed by radical–molecule, molecular growth pathways.

Zone 4, the gas quench or “cool zone,” exists outside the combustion chamber and can be characterized by either gradual or rapid quenching of the gas temperature. Residence times can be very long, up to 10 s, and oxygen concentrations can vary from oxygen depleted, due to combustion in upstream zones, to very oxygen rich if air in-leakage occurs. The “cool zone” has not been well studied, but may in fact be responsible for a significant fraction of observed pollutant emissions from many types of combustors [20]. Partially oxidized products such as formaldehyde, chloroformaldehyde, and phosgene can form by radical–oxygen association reactions [21]. Nitrated products can also form via radical–molecule addition reactions involving NO_x species generated in the upstream, higher temperature zones [22].

Zone 5, the surface catalysis zone, exists at largely the same longitudinal point in the gas flow-path as the

“cool zone” (zone 4). However, the distinguishing aspect of this zone that is fundamentally different from zones 1 to 4 is that one must now consider the effects of surfaces. Reaction times for gas–surface reactions can be the same as in zone 4 for entrained particulate; however, reaction times involving deposited particulate matter can be orders of magnitude longer before re-entrainment occurs and even longer if fixed surfaces are involved in the reaction. The temperature range in this zone is about 423–773 K. Below 423 K, sufficient energy is not available for activation of chemisorptions, and above 773 K most reactants are desorbed. This zone has been the subject of somewhat limited study, with the surface-mediated formation of polychlorinated dibenzo-p-dioxins (PCDDs) and polychlorinated dibenzofurans (PCDFs) receiving the only serious attention [23–27]. However, many more pollutants can potentially form as a result of surface catalysis via pathways including *de novo* synthesis of simple and more complex CHCs, partial oxidation of hydrocarbons and CHCs to form carbonyls, alcohols, organic acids, epoxides, etc., and nitration-type reactions to form a range of nitrogen-containing

products [24, 28, 29]. However, most of the reactions necessary to form these products likely require a transition metal catalyst in addition to an adsorption site [26, 30–34].

Table 2 summarizes the conditions and types of reactions (in order of importance) expected in each of five reaction zones. In the most general sense, the mechanisms of pollutant formation and destruction are expected to be relatively consistent within a zone. This “Zone Model” has the advantage that one can classify the types of reactions occurring within a given zone and consider their impact on emissions in more detail [35]. Table 2 also shows how the conditions within each zone may be distinctly different, resulting in different mechanisms of (POHC) destruction and (PIC) formation being dominant in each zone. Four classes of mechanisms of POHC destruction have been defined: (1) concerted molecular elimination, (2) bond fission, (3) bimolecular radical attack, and (4) surface-catalyzed decomposition. Four classes of

mechanisms of PIC formation are included: (1) concerted molecular elimination of stable molecules, (2) complex radical–molecule pathways, (3) radical–radical (atom) and radical–molecule association reactions, and (4) surface-catalyzed synthesis.

Although there will be some contribution of each mechanism in zones 1–3, chemical kinetic theory suggests that mechanism 1 will dominate in the pre-flame zone, mechanisms 2 and 3 will dominate in the flame zone, and mechanisms 1 and 3 will dominate in the post-flame thermal zone. The gas quench and surface catalyst zones are somewhat different, due to the much lower temperatures and longer residences describing these zones. Mechanism 1 and 3 will dominate in the gas quench zone, and mechanism 4, by definition, will dominate in the surface catalyst zone. The nature and yield of by-product emissions will depend upon the conditions in a particular incinerator as well as the elemental and molecular chemical composition of the waste/fuel feed.

Non-flame Incineration. Table 2 Dominant mechanisms of pollutant destruction and formation^a (From [35])

	Zone	Reaction conditions	Decomposition mechanisms ^b	Formation mechanisms ^b
1	Pre-flame	T = 473–1273 K	i	i
		t _r << 1 s	ii	ii
		[O ₂] ~ 50% excess air	iii	iii
2	Flame	T = 1273–2073 K	iii	ii
		t _r ≤ 0.01 s	ii	i
		[O ₂] ~ 50% excess air	i	iii
3	High-temperature thermal (non-flame)	T = 873–1373 K	i	iii
		t _r = 1–10 s	ii	ii
		[O ₂] ~ 50–100% excess air	iii	i
4	Gas quench (cool zone)	T = 353–873 K	i	iii
		t _r range: 1–10 s	ii	
		[O ₂] ~ 3–9%		
5	Surface catalyst	T = 423–773 K	iv	iv
		t _r range: 10 s–10 min		
		[O ₂] ~ 3–9%		

^aDecomposition mechanism: i. concerted molecular elimination; ii. bond fission; iii. bimolecular radical attack; iv. surface-catalyzed decomposition. Formation mechanism: i. concerted molecular elimination; ii. complex radical–molecule pathways; iii. recombination and association reactions; iv. surface-catalyzed synthesis

^bIn decreasing order of importance

PIC Emissions from Full-Scale Facilities

Tables 3 and 4 present specific PIC emissions data calculated from raw data available from field testing of nine hazardous waste incinerators [7, 36]. For each compound, frequency of occurrence, arithmetic mean emission rate, and geometric mean emission rate are presented. In calculation of the arithmetic and geometric mean emission rates, only values greater than analytical detection limits were considered, i.e., detection limit values were not assigned a numerical value.

Inspection of the full-scale incinerator data indicates that chloroform, benzene, tetrachloroethene,

toluene, carbon tetrachloride, methylene chloride, 1,1,1-trichloroethane, chlorobenzene, and naphthalene are the most prevalent PICs, emitted in at least 75% of all tests. It should be noted that occurrence frequency in this context excludes the presence of the compound in the waste feed as a POHC. Of these compounds, chloroform, methylene chloride, benzene, toluene, and naphthalene exhibited largest emission rates ($>1,000 \mu\text{g}/\text{min}$). The greater values of arithmetic versus geometric mean emission rates for many of these compounds indicate that absence of a normal distribution of emission data, i.e., the distribution is skewed

Non-flame Incineration. Table 3 Frequently occurring volatile PICs in full-scale research tests

Compound	Frequency of occurrence ^a	Arithmetic mean emission rate ^b ($\mu\text{g}/\text{min}$)	Geometric mean emission rate ^c ($\mu\text{g}/\text{min}$)	Formation mechanism
Benzene	6/6	38396.8	6756.5	Radical–molecule
Chloroform	6/6	31616.3	2549.5	Association
Toluene	3/3	4198.7	1981.7	Radical–molecule
Carbon tetrachloride	1/1	126.3	120.4	Association
Tetrachloroethene	4/4	378.0	107.1	Radical–molecule
Chlorobenzene	4/5	674.0	494.8	Radical–molecule
1,1,1-Trichloroethane	4/5	520.3	84.3	Association
Methylene chloride	3/4	11707.7	1623.0	Radical–molecule
Bromodichloromethane	6/9	4794.7	2295.5	Association
Dibromochloromethane	6/9	1592.8	1059.6	Association
Bromoform	5/9	8254.8	2419.3	Association
Trichloroethene	1/2	762.0	721.1	Radical–molecule
Bromochloromethane	3/9	7334.0	1733.5	Association
Chloromethane	2/8	197481.0	161375.0	Radical–molecule
Methyl ethyl ketone	1/7	172.0	169.0	Radical–molecule
Methylene bromide	1/8	12800.0	12554.0	Association
Bromomethane	1/8	2300.0	2291.3	Association
Chloroethene	1/9	2405.0	2144.0	Molecule elim.
1,2-Dichloroethane	1/9	1946.0	1510.0	Association

^aDefined as number of occurrences as PIC divided by total number of possible occurrences as PIC; denominator excludes presence in waste feed

^bArithmetic mean emission rate for all sites; individual data (per site) consist of an arithmetic mean based on the number of measurements obtained above detection limits

^cGeometric mean emission rate for all sites; individual data (per site) consist of geometric mean based on the number of measurements obtained above detection limits

Non-flame Incineration. Table 4 Frequently occurring semivolatile PICs in full-scale research tests

Compound	Frequency of occurrence ^a	Arithmetic mean emission rate ^b (μg/min)	Geometric mean emission rate ^c (μg/min)	Formation mechanism
Naphthalene	3/4	2764.3	1461.2	Radical–molecule
Phenol	2/6	4326.5	1633.0	Association
2-Nitrophenol	2/8	4971.5	3398.5	Association
2,4,6-Trichlorophenol	1/8	5867.0	5867.0	Association
2,4-Dimethylphenol	1/8	1600.0	1600.0	Association
Pyrene	1/8	1100.0	1100.0	Radical–molecule
2,6-Toluene diisocyanate	1/8	200.0	200.0	Radical–molecule
Hexachlorobenzene	1/8	24.0	24.0	Radical–molecule
2-Chlorophenol	1/8	479.0	479.0	Association
Pentachlorophenol	1/8	340.0	340.0	Association
Fluoranthene	1/8	700.0	700.0	Radical–molecule

^aDefined as number of occurrences as PIC divided by total number of possible occurrences as PIC; denominator excludes presence in waste feed

^bArithmetic mean emission rate for all sites; individual data (per site) consist of an arithmetic mean based on the number of measurements obtained above detection limits

^cGeometric mean emission rate for all sites; individual data (per site) consist of geometric mean based on the number of measurements obtained above detection limits

towards larger emission rates. Considering the scatter in the data, the geometric mean emission rate is a more representative measure of the actual emission rate for these compounds.

Further inspection of Table 3 indicates that chloromethane, methylene bromide, bromomethane, bromoform, bromodichloromethane, dibromochloromethane, and bromochloromethane also exhibited emission rates comparable to or greater than the more frequently observed compounds. There are at least three potential causes for the intermittent emission of these brominated halomethane PICs: (1) they are formed (via true PIC formation) from bromine-containing precursors in the waste or fuel which were not reliably analyzed for by full-scale sampling/analytical techniques, (2) they are contained in nearly all waste streams yet not reliably analyzed by full-scale analytical techniques, and (3) they are formed by volatile compound stripping from makeup scrubber water. The relative contribution of these sources cannot be ascertained with the available data. However, the fragile nature of these compounds suggests they cannot survive a well-operated incinerator. Thus, their

presence in the waste feed in the same chemical form in which they are omitted is unlikely. “Cool zone” formation or stripping from the scrubber water is thus more likely sources.

The Zone Theory of PIC formation delineates three specific PIC formation mechanisms based on an analysis of laboratory-scale PIC formation and destruction behavior. These mechanisms are also listed in Tables 3 and 4 for the specific PICs identified from the incinerator evaluation. For the types of PICs observed, emission mechanisms are divided into three categories: molecular elimination reactions, radical–molecule reactions, and radical association reactions. The relative contributions of these three mechanisms to the observed PIC emissions depend on the relative importance of the following parameters: (1) frequency of occurrence and (2) geometric mean emission rate. Weighting the former as more significant suggests that 67% of the PICs are emitted by a radical–molecule mechanism. However, based on geometric mean emission rate (>1,000 μg/min), the data indicate that both radical–molecule and radical association mechanisms are of equal significant in producing PIC emissions.

An equal weighting of these two parameters indicates that 67% of the PICs are emitted by radical–molecule mechanisms (weighted geometric mean emission rate $>1,000$ $\mu\text{g}/\text{min}$). The above analysis excludes brominated halomethane compounds because their mechanism of formation may be partially due to stripping of scrubber makeup water. Observed phthalate compounds have also been excluded because their mechanism of formation are at least partially due to ubiquitous sources, e.g., outgassing of phthalates in plasticizers used in analytical laboratories.

In summary, the field data base indicates that the most prevalent class of PICs emitted was the chlorinated methanes. Benzene, toluene, and naphthalene represented the most prevalent aromatic species observed. PICs of secondary prevalence, yet still observed in at least 50% of possible cases, included the chlorinated ethenes and the chlorinated ethanes. Another major observation from the analysis of the specific PIC emissions is related to their mechanism of formation. The data from all types of thermal disposal facilities indicate that radical–molecule reactions and radical association reactions are by far the dominant emission mechanisms. This conclusion is reached independent of the relative weighting of PIC classification parameters. Analysis of specific PICs generated by these two emission mechanisms indicated that radical association reactions varied in significance from $\sim 30\%$ to as high as 50% , depending on the emission source. As discussed in the Zone Theory of PIC Formation, radical–molecule and radical–radical association reactions are the dominant chemical reactions occurring in the post-flame and cool zones of incineration systems. Although the available PIC emissions data base is very limited, the mechanistic implications of this analysis suggest that the cooler region downstream of the flame zone may be responsible for a significant fraction of the observed emissions. The degree to which these types of reactions contribute to the observed emissions is dependent on several facility-dependent factors, the most important perhaps being post-flame zone waste/air mixing and cool zone quenching rates.

Examination of full-scale PIC measurements for more than a dozen incinerator compliance and test burns has resulted in the observation that there are several common or “generic” PICs that tend to be emitted regardless of the waste being fed. With

the exception of methyl chloride and vinyl chloride, these are all hydrocarbons that likely result from both waste and fuel combustion. Surprisingly, the yields of these PICs do not seem to vary greatly from incinerator to incinerator. Thus, it has been possible to develop approximate PIC emission factors. Table 5 lists these values in μg of PIC per kg of waste feed. Although one still expects significant variations of actual emission rates from those calculated based on these values, experience has shown them to be accurate to within an order of magnitude. Furthermore, the relative emission rates of each PIC have been found to be quite consistent with these emission factors.

Formation of Chlorinated Polycyclic Aromatic Hydrocarbons

In this section, laboratory studies of the molecular growth reactions of CHCs are reviewed. Mechanisms and rate parameters associated with the formation of chlorinated aromatic compounds are emphasized.

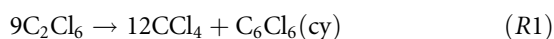
Non-flame Incineration. Table 5 Approximate emission factors for consistently observed PICs

Compound	Emission factor ($\mu\text{g}/\text{kg}$)
Biphenyl	63
Benzene	31
Toluene	30
Naphthalene	22
Phenol	8.7
Benzaldehyde	5.5
Ethyl benzene	5.1
Xylene	4.3
Benzoic acid	2.7
Benzofuran	5.0
Methyl chloride	5.0
Vinyl chloride	0.88
Acenaphthalene	0.50
Acenaphthene	0.50
Fluoroanthene	0.50
Phenanthrene	0.20
Dibenzofuran	0.01

From an environmental perspective, the molecular growth of CHCs is significant because by-products of these reactions include chlorinated aromatic and chlorinated polycyclic aromatic hydrocarbons (ClPAH). The potential emission of these products into the environment is of concern due to their potential health effects [37].

Gas-Phase Molecular Growth Reactions

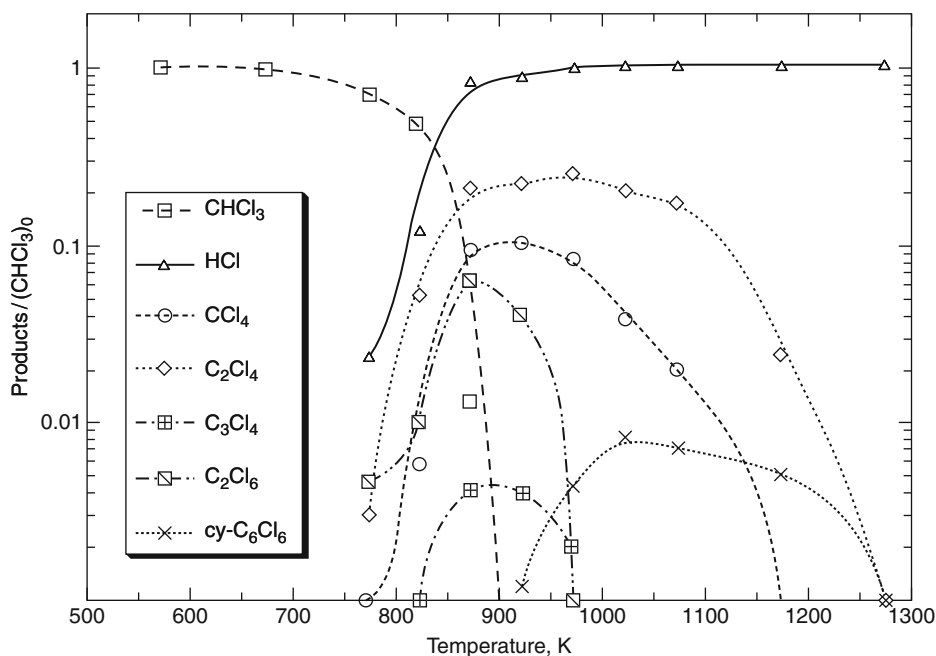
Nearly 50 years ago, Aubrey and van Wazer [38] showed empirically the existence of an equilibrium between chloro-aliphatic and aromatic compounds:



There were several further investigations on thermal reactivity of chloro-aliphatic and aromatic compounds between 1960 and 1980. Most of these early studies were performed using scaled ampoule techniques. For a review, see Choudhry and Hutzinger [39]. Since 1980, modern methods of elucidating reaction mechanisms using hyphenated analytical

techniques (typically gas chromatography coupled to specific detectors) interfaced directly to special flow reactors, and drop tube reactors have been used to unravel the complexity of non-flame thermal reactions of chloro-organic compounds under both oxidative and pyrolytic conditions. These more recent studies are updated here.

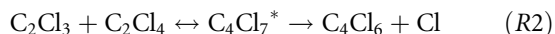
The oxidative pyrolysis of methylene chloride (CH_2Cl_2), chloroform (CHCl_3), and carbon tetrachloride (CCl_4) has been investigated by Tirey et al. [40] and Taylor et al. [41] using a fused silica, tubular flow reactor coupled to gas chromatographic–mass spectrometric (GC–MS) detection. This study first illustrated the dominance of pyrolytic reaction behavior leading to the formation of several thermally stable reaction by-products and the observation of chlorinated molecular growth reactions resulting in the formation of hexachlorobenzene from simple chlorinated methane precursors. The facile formation of C_6Cl_6 (cy) from the oxidative pyrolysis of CHCl_3 is illustrated in Fig. 2. Key elementary reactions leading to the formation of C_6Cl_6 (cy) are presented and discussed by Taylor and Dellinger [42].



Non-flame Incineration. Figure 2

Product distributions from the oxidative pyrolysis of CHCl_3 . CHCl_3/O_2 equiv. ratio = 3.0, $[\text{CHCl}_3]_0 = 2.7 \times 10^{-5}$ mol/L, $t_r = 2.0$ s, $P = 1.15$ atm. Curves drawn through data based on smoothed interpolated fit (From [41])

The formation of higher molecular weight compounds from the pyrolysis and combustion of CHCs has been observed for the following precursors: CHCl_3 [40, 41, 43], CCl_4 [40, 41, 43], dichloroacetylene (C_2Cl_2) [44], trichloroethene (C_2HCl_3), [45–47], tetrachloroethene (C_2Cl_4) [48–50], hexachloropropene (C_3Cl_6) [51], and 1,3-hexachlorobutadiene (C_4Cl_6) [52, 53]. Tirey et al. [49] proposed a mechanism including the formation of the following intermediate species: C_2Cl_2 , C_2Cl_3 , C_2Cl_4 , C_2HCl_3 , C_4Cl_4 , C_4Cl_5 , and C_4Cl_6 that required further experimental and theoretical analysis. The mechanism was based on mechanisms proposed to account for observed products from the pyrolysis and combustion of unsaturated aliphatic hydrocarbons [54, 55] and methyl chloride [56]. However, based on the absence of production of chlorinated aromatics from CH_3Cl and CH_2Cl_2 as compared with CHCl_3 and CCl_4 , Dellinger and Taylor [42] proposed that chlorine facilitates condensation reactions resulting in the formation of chlorinated aromatic species. They suggested that the tendency to form high molecular weight species is attributable to the propensity for chlorinated radicals to undergo reversible, addition/elimination (displacement)-type molecular growth reactions:



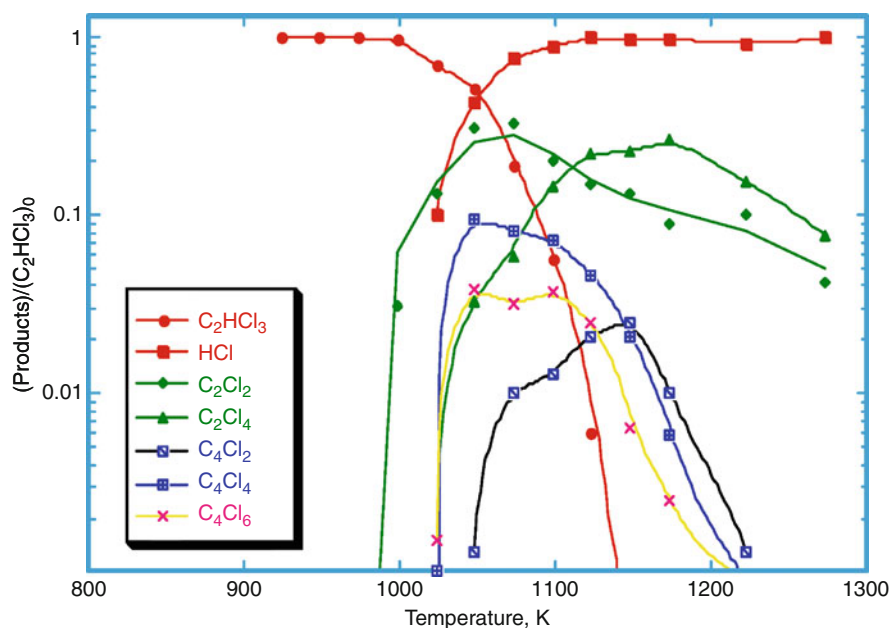
Chlorine substitution can play a major role in the rate of these molecular growth reactions through inductive destabilization of the initially formed adduct (*) and through rapid elimination of chlorine. Chlorine elimination reduces the probability of the reverse decomposition to reactants. The lower C–Cl bond energies in CHCs (vs. C–H bond energies in hydrocarbons) favor these types of molecular growth reactions. Tirey et al. [49] and Dellinger and Taylor [57] presented Quantum RRR calculations illustrating that chemically activated Cl displacement from olefinic and acetylenic species by olefinic radicals is favored for CHCs as compared with the similar displacement of H atoms from hydrocarbons. They proposed that this is in part responsible for the high yields of chlorinated aromatics for the more chlorinated of these species. The following paragraphs summarize recent studies of the pyrolysis and combustion of C_2 unsaturated CHCs and the development of core reaction models

describing molecular growth leading to the formation of chlorinated aromatic species.

Measurements of condensed-phase products of C_2HCl_3 pyrolysis in a drop-tube reactor obtained over a temperature range of 1,100–1,500 K were reported by Mulholland et al. [45]. A predominance of perchloroaromatics was found, in sharp contrast to the largely unsubstituted PAHs obtained from the pyrolysis of a mixture of trichloroethylene and toluene under similar conditions. Detailed characterization of product tars from the pyrolysis of C_2HCl_3 using GC–MS analysis revealed the following molecular growth products: C_4Cl_6 , C_6HCl_5 , C_6Cl_6 , $\text{C}_8\text{H}_4\text{Cl}_6$, C_8Cl_6 , C_7HCl_7 , C_8Cl_8 , $\text{C}_{10}\text{H}_4\text{Cl}_6$, C_9Cl_8 , C_{10}Cl_8 , $\text{C}_{10}\text{HCl}_7$, C_{12}Cl_8 , $\text{C}_{12}\text{Cl}_{10}$, and C_{14}Cl_8 . A reaction mechanism beginning with C_2Cl_2 formation from C_2HCl_3 via concerted HCl elimination was proposed. Aromatic ring formation and growth were proposed to occur by successive additions of C_2Cl_2 to C_2HCl_3 , in a manner similar to that proposed by Dellinger and Taylor [57].

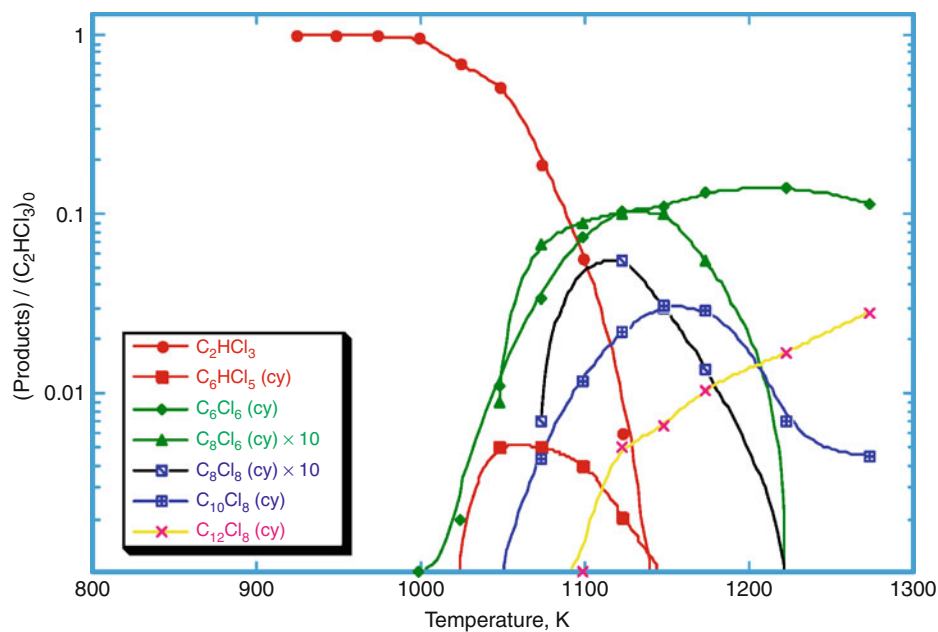
The oxygen-free pyrolysis of C_2HCl_3 was investigated by Taylor et al. [46, 47] using a fused silica, tubular flow reactor coupled to in-line GC–MS detection. The major degradation by-products and molecular growth by-products are summarized in Figs. 3 and 4, respectively. Pronounced molecular growth was observed above 1,000 K as evidenced by the formation of C_2Cl_4 , C_4Cl_6 , and C_6Cl_6 (cy) as major (> 5 mol %) products and dichlorodiacetylene (C_4Cl_2), pentachlorobenzene (C_6HCl_5 (cy)), hexachloroethynylbenzene (C_8Cl_6 , (cy)), octachlorostyrene (C_8Cl_8 (cy)), octachloronaphthalene (C_{10}Cl_8 (cy)), and C_{12}Cl_8 (cy) as minor (< 5 mol %) products. A detailed reaction kinetic model was developed describing molecular growth up to the formation of C_8Cl_6 (cy) and C_8Cl_8 (cy).

A core pyrolysis model that accounts for the major products up to C_8Cl_6 (cy) and C_8Cl_8 (cy) is depicted in Table 6. Sensitivity and production rate analyses indicated that C–Cl bond fission is the dominant initiation step. Cl atoms abstract H from C_2HCl_3 yielding C_2Cl_3 and HCl. Trichlorovinyl decomposes by C–Cl scission to form C_2Cl_2 or reacts with Cl_2 to give C_2Cl_4 . Alternatively, C_2Cl_3 can add to C_2HCl_3 , C_2Cl_2 , or C_2Cl_4 via Cl displacement to give C_4HCl_5 , C_4Cl_4 , or C_4Cl_6 , respectively. Reaction of C_4Cl_5 and C_4Cl_3



Non-flame Incineration. Figure 3

Major products from the oxygen-free pyrolysis of C_2HCl_3 . $[C_2HCl_3]_0 = 5.4 \times 10^{-5}$ mol/L, $t_r = 1.7$ s, $P = 1.83$ atm



Non-flame Incineration. Figure 4

Chlorinated aromatic products from the oxygen-free pyrolysis of C_2HCl_3 . $[C_2HCl_3]_0 = 5.4 \times 10^{-5}$ mol/L, $t_r = 1.7$ s, $P = 1.83$ atm

Non-flame Incineration. Table 6 C_2HCl_3 Core pyrolysis and molecular growth mechanism (From [47])

Reaction	log k (cm ³ /mol/s/kcal)
$C_2HCl_3 \rightarrow C_2HCl_2 + Cl$	15.0-91.0/2.303 RT
$C_2HCl_3 + Cl \rightarrow C_2Cl_3 + HCl$	12.3-9.5/2.303 RT
$C_2Cl_3 + M \rightarrow C_2Cl_2 + Cl + M$	15.0-32.8/2.303 RT
$C_2Cl_3 + Cl_2 \rightarrow C_2Cl_4 + Cl$	12.5-4.8/2.303 RT
$C_2Cl_3 + C_2HCl_3 \rightarrow C_4HCl_6$	11.0-7.0/2.303 RT
$C_4HCl_6 \rightarrow C_4HCl_5 + Cl$	13.0-18.0/2.303 RT
$C_4HCl_5 + Cl \rightarrow C_4Cl_5 (i) + HCl$	12.3-5.0/2.303 RT
$C_4Cl_5 (i) \rightarrow C_4Cl_4 + Cl$	13.4-43.9/2.303 RT
$C_2Cl_3 + C_2Cl_2 \rightarrow C_4Cl_5 (n)$	12.0-6.0/2.303 RT
$C_4Cl_5 (n) \rightarrow C_4Cl_4 + Cl$	13.4-34.1/2.303 RT
$C_4Cl_4 \rightarrow C_4Cl_3 (i) + Cl$	16.0-68.4/2.303 RT
$C_4Cl_4 \rightarrow C_4Cl_3 (n) + Cl$	16.0-84.0/2.303 RT
$C_4Cl_3 (n) + C_2Cl_2 \rightarrow C_6Cl_5 (l)$	11.7-5.0/2.303 RT
$C_6Cl_5 (l) \rightarrow C_6Cl_5 (cy)$	10.0
$C_6Cl_5 (cy) + Cl \rightarrow C_6Cl_6 (cy)$	12.6 + 1.2/2.303 RT
$C_4Cl_5 (n) + C_2Cl_2 \rightarrow C_6Cl_7 (l)$	12.5-5.0/2.303 RT
$C_6Cl_7 (l) \rightarrow C_6Cl_6 (cy) + Cl$	10.0
$C_6Cl_5 (cy) + C_2Cl_2 \rightarrow C_8Cl_6 (cy) + Cl$	11.7-5.0/2.303 RT
$C_6Cl_5 (cy) + C_2Cl_4 \rightarrow C_8Cl_8 (cy) + Cl$	11.7-5.0/2.303 RT

radicals with C_2Cl_2 are the dominant pathways to formation of C_6Cl_6 (cy). Reaction of C_6Cl_5 (cy) with C_2Cl_4 and C_2Cl_3 with C_6Cl_6 (cy) contribute nearly equally to C_8Cl_8 (cy) formation while the reaction of C_2Cl_2 with C_6Cl_5 (cy) dominates C_8Cl_6 (cy) formation. The conversion of C_2Cl_3 into C_2Cl_4 inhibits both the formation of larger molecular weight species and the destruction of C_2HCl_3 .

Except for the very lowest temperatures, model calculations indicate that the C_2Cl_3 radical chain is the dominant conversion cycle for C_2HCl_3 under pyrolytic conditions. Rapid abstraction of H from C_2HCl_3 by Cl atoms removes most H-containing species from the reaction system before C_4 species can be formed. Hydrogen-containing C_2 or C_4 species that form at

trace levels are rapidly destroyed by additional H abstraction by Cl. When a hydrogen-rich fuel is added, there is competition for the available Cl atoms, and H-containing CHCs would be expected to survive and result in the formation of a variety of H-containing molecules. This is in agreement with observations of others working on mixed H and Cl systems [45].

Experimental and modeling studies were also carried out by Taylor et al. on other unsaturated CHCs including C_2Cl_4 [50], C_3Cl_6 [51], and C_4Cl_6 [52] under conditions very similar to the C_2HCl_3 experiments. Similar molecular growth reaction products were observed. Attempts to model this data using similar reaction pathways developed for the C_2HCl_3 degradation products were generally successful with some exceptions.

The implication of the results of the experimental and modeling effort on CHC pyrolysis is the manner in which weaker C–Cl bonds (vs. C–H bonds) facilitate certain reactions. Chlorine atom displacement, internal chlorine atom abstractions, and chlorine migrations can all facilitate molecular growth. Formation of perchlorinated aromatic species and ClPAH that are resistant to oxidation may actually be more kinetically favored than formation of PAH from the corresponding hydrocarbons.

Formation of Polychlorinated Dibenzo-P-Dioxins and Furans

In the previous section, it was demonstrated that complex aromatic CHCs readily form from aliphatic CHCs at elevated temperatures. Combustion processes are one of the major sources of PCDD/F in the environment ([58, 59]). The concerns regarding environmental exposure to PCDD/F are well known, see, for example, the US-EPA Dioxin Reassessment by the National Academy of Sciences [60].

Gas-phase molecular growth mechanisms of CHCs leading to the formation of PCDD/F have been unable to account for the PCDD/F concentrations observed in the flue gas of full-scale combustors [61, 62]. As a result, over the past two decades, research on the mechanism of formation of PCDD/F in combustion sources has focused on surface-mediated pathways [23–27, 63, 64]. In this section, we review that state of

knowledge of both low-temperature, surface-mediated formation of PCDD/F and recent efforts to develop a kinetic model that can predict the higher temperature, surface-mediated formation of PCDD/F from chlorophenol precursors.

Current Status of Surface-Mediated (Low Temperature) Studies of PCDD/F Formation

Seventy percent of emissions of PCDD/F from combustion and thermal processes have been estimated to occur via transition metal-mediated reactions, in the post-flame, cool zone [65–67]. Two surface-mediated formation pathways have been proposed: (1) the *de novo* pathway in which elemental carbon and transition metals in combustion-generated particles react with oxygen and chlorine to liberate PCDD/F from the carbon matrix and (2) the precursor pathway in which organic species such as chlorophenols react on a transition metal surface to form PCDD/F [23–27, 33–35, 39, 61, 63–70].

In the *de novo* pathway, it has recently been proposed that nonchlorinated dibenzofuran and dibenzo-*p*-dioxins are first formed from the carbon and then chlorinated to form PCDFs and lesser amounts of PCDDs [71–75]. Research on the precursor pathway has demonstrated that chlorophenols are readily converted to PCDDs and lesser amounts of PCDFs by reaction over a copper oxide surface [23, 76, 77]. The high yields of PCDFs from the *de novo* process and the high yields of PCDDs from the precursor pathway have led to the suggestion that PCDFs are formed primarily from elemental carbon (or PAHs), and PCDDs are formed primarily from chlorophenols in full-scale incinerators.

In apparent contradiction to the chlorophenol hypothesis, correlations between various incineration parameters and PCDD/F emission concentrations indicated chlorobenzene concentrations, rather than chlorophenol concentrations, were the best predictors of total PCDD/F emissions [24, 27, 67, 78–80]. This finding is complicated by the fact that chlorophenols are difficult to analyze due to their propensity to strongly bind with the surface of particulate matter. However, there is a lack of experimental evidence of formation of PCDD/Fs from chlorobenzenes [81, 82]. In spite of this, it is frequently assumed that

chlorobenzenes readily form PCDD/Fs in full-scale systems [24, 27].

Although research has demonstrated copper oxide-mediated reactions can convert simple hydrocarbons such as ethylene and acetylene to chlorobenzenes and CIPAHs [26, 32, 44, 83, 84], chlorophenols have not been detected as intermediates in these experiments. Chlorophenols do not form from chlorobenzenes during gas-phase pyrolysis, as chlorobenzenes undergo ring rupture with a small fraction converting to other substituted aromatic compounds under oxidative conditions [85, 86].

A recent study has attempted to determine if chlorobenzenes can be converted to PCDD/Fs in sufficient yields to make a significant contribution to their emissions from full-scale facilities. Nganai et al. [34] investigated the potential contribution of chlorinated benzenes to the formation of PCDD/Fs using 1,2-dichlorobenzene as a surrogate for reactions of other chlorinated benzenes and CuO/silica (3 wt.% Cu) as a surrogate for fly ash. Results were similar for oxidative and pyrolytic conditions with a slight increase in more chlorinated products under oxidative conditions. Reaction products included chlorobenzene, polychlorinated benzenes, phenol, 2-chlorophenol (2-CP), dichlorophenols, and trichlorophenols with yields ranging from 0.01% to 2% for the phenols and from 0.01% to 10% for chlorinated benzenes. 4,6-Dichlorodibenzofuran (4,6-DCDF) and dibenzofuran (DF) were observed in maximum yields of 0.2% and 0.5%, respectively, under pyrolytic conditions and 0.1% and 0.3%, respectively, under oxidative conditions. In previous studies of the pyrolysis of 2-CP under identical conditions [33], 4,6-DCDF and dibenzo-*p*-dioxin (DD) were observed with maximum yields of ~0.2% and ~0.1%, respectively, along with trace quantities of 1-chlorodibenzo-*p*-dioxin (1-CDD). Under oxidative conditions, DD, 1-CDD, and 4,6-DCDF were observed with maximum yields of 0.07%, 0.3%, and 0.1%, respectively. When combined with the fact that measured concentrations of chlorobenzenes are often 10–100 times higher than chlorophenols in full-scale combustion systems, these studies suggest that surface-mediated reactions of chlorinated benzenes can be a significant source of PCDD/F emissions.

Current Status of Modeling of Higher Temperature PCDD Formation from Chlorophenol Precursors

Shaub and Tsang [61, 87] initially explored this issue through the development of a simple 13-step, gas-phase reaction mechanism and a separate heterogeneous catalytic model of dioxin formation and concluded that heterogeneous rates were much faster than homogeneous rates. Furthermore, they concluded that the more rapid heterogeneous rates were necessary to account for full-scale municipal waste incinerator data. Sidhu et al. [88] challenged the kinetic rates used in the gas-phase dioxin model and concluded, based on experimental measurements of PCDD/F formation from 2,4,6-trichlorophenol, that the estimated rate for pre-dioxin formation from trichlorophenol precursors was underestimated by Shaub and Tsang [61]. Grotheer and Louw [89] conducted an experimental study to verify the hypothesis of Sidhu et al. [88] and concluded that the homogeneous rates in the reaction mechanism proposed by Shaub and Tsang [61] were indeed correct and that their experiments and those by Sidhu et al. could be interpreted in terms of a different gas-phase model incorporating phenoxy radical recombination as the key step in dioxin formation.

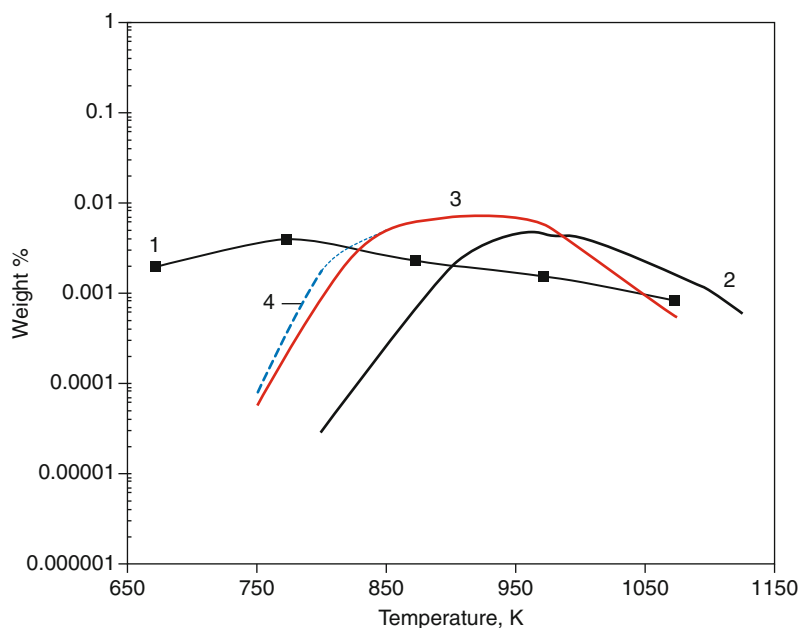
With the realization that phenoxy radicals are exceptionally unreactive with oxygen, a modified gas-phase PCDD model has recently been proposed that could explain the formation of PCDD by purely gas-phase reactions of phenoxy radicals [90, 91]. Published models have assessed the relative importance of the role of labile chain carriers including H, Cl, O, OH, HO₂, and alkyl radicals (R) [90–93]. In the recently published modeling of Khachatryan et al., OH proved to be the dominant chain carrier [90, 92, 93]. However, other recent publications suggest that Cl is a highly active chain carrier in the formation of PCDD and chlorinated aromatic pollutants as well as being a chlorinating agent [91, 94]. In these models, the main channel for the decomposition of chlorinated phenoxy radicals was expulsion of CO (the dominant oxygen-containing product) and formation of chlorinated cyclopentadienyl radical, C₅H₂Cl₃ [90, 92, 93]. For the sake of simplicity, the further decomposition of C₅H₂Cl₃ was not considered, which leads to a low steady-state concentration of Cl versus OH [91]. To more fully assess the role of Cl in overall chain

processes over wide temperature ranges, it is important to release Cl atoms from C₅H₂Cl₃ (as one of the major products) using reasonable decomposition pathways and associated reaction kinetics.

The roles of OH and Cl in the formation of 1,3,6,8-tetrachlorodibenzo-p-dioxin (1,3,6,8-TCDD) from 2,4,6-trichlorophenol (2,4,6-TCP) were recently assessed by Khachatryan et al. [95] using an expanded PCDD model that includes several new reaction schemes that were added to the original model [90]. The final model of 132 reactions (given in Scheme 1 of the Supporting Information of Khachatryan et al. [95]) includes updated sub-models for TCDD formation; hydrogen oxidation; hexane oxidation; recombination/decomposition of C₅H₂Cl₃ and release of Cl; generation of Cl through reactions of Cl₂ with OH, H, and O; additional reactions for release of Cl [96]; and an additional four-step surface-mediated generation of 1,3,6,8-TCDD sub-model (see Table 4 of [95]).

The modifications to the model resulted in improved agreement with experimental results [97] below 900 K as illustrated in Fig. 5. The maximum deviation of calculated and experimental yields for 1,3,6,8-TCDD was less than factor of 4 in the temperature region between 823 and 1,073 K, and the maximum experimental yield of 0.004 wt% at 773 K was very close to the calculated yield 0.006% at 948 K despite an approximate 150 K shift in the maximum. Considering the large uncertainties in some of the thermochemistry and kinetic parameters of key reactions, the agreement is quite remarkable.

The experimentally observed yields of 1,3,6,8-TCDD between 648 and 798 K [97] are still higher than calculated. However, recent publications on the pyrolysis of 2-CP and 2,4,6-TCP reported lower PCDD yields in this temperature regime [98, 99]. PCDDs were not detected below 748 K from the gas-phase pyrolysis of 2-CP [98]. The oxidative pyrolysis of 2,4,6-TCP resulted in the formation of 1,3,6,8- and 1,3,7,9-TCDD between 773 and 1,053 K with a maximum at 898 K [99]. The results of this model agree reasonably well with these experimental results over the temperature range of 823–1,073 K. Khachatryan et al. [95] included very favorable rates for silica-catalyzed reactions in the revised model. It is thus possible that the higher yields of 1,3,6,8-TCDD



Non-flame Incineration. Figure 5

Comparison of gas-phase experimental and modeling results for 1,3,6,8-TCDD formation. *Curve 1*: experimental results [97]. *Curve 2*: gas-phase model [90]. *Curve 3*: revised gas-phase model – 132 gas-phase reactions [95]. *Curve 4*: combined gas-phase and surface reaction model – 136 reactions including four reaction surface sub-model [95] (From [95])

between 648 and 848 K for the reference experimental study were due to wall deposits of carbon or other impurities [27] that are not accounted for in the current model.

Modeling results indicate that the role of CI is limited during pyrolysis of highly chlorinated phenols because CI is incorporated into the reaction products. Although the OH to CI ratio is <1 in the equilibrium calculations, it is always >1 for the reaction kinetic schemes with 136 reactions considered here. This suggests that OH is the dominant chain carrier below 1,048 K, and CI only participates as a secondary chain carrier at higher temperatures.

Future Directions

Thermal Stability–Based Incinerability Ranking

The thermal stability–based incinerability ranking does not currently have experimental indices for highly chlorinated aromatic compounds with expected high thermal stability and high toxicity. These compounds would include specific isomers of the family of PCBs and PCDD/Fs, in addition to specific ClPAHs

of suspected high toxicity. The addition of these compounds would provide expanded utility of this important ranking for full-scale trial burns.

Formation of Products of Incomplete Combustion

Field tests of large scale incinerators yield emissions data that are not readily explainable using flame or post-flame chemistry, exclusively. Many or most of the observed emissions are products that are thermally fragile and are not expected to survive the flame zones of a thermal destruction facility. Thus, “cool zone” chemistry is implicated. As a result of the need to reduce PCDD/F emissions, surface-catalyzed chlorination reactions under “cool zone” conditions have received the most attention to date. This research has been very successfully translated to full-scale thermal treatment facilities where overall PCDD/F emissions have decreased dramatically over the past 20 years. However, recent characterization of combustor effluents using a technique known as multidimensional gas-chromatography–mass spectrometry (MDGC–MS) has shown that a large number of oxygenated PICs are

emitted [100]. These are thermally fragile species that are most likely formed in the “cool zone” of the incinerator. Surface-catalyzed partial oxidation is suspected as the route of formation. If this is true, then these species are likely to be emitted from virtually any combustion system. Since some of the observed pollutants are known endocrine disrupting chemicals (EDCs), their route of formation and their relationship to formation of other pollutant emissions from thermal destruction facilities is meritorious of future study.

CHC Molecular Growth and PCDD Modeling

The limitation in more accurate CHC molecular growth models is the lack of experimental elementary rate measurements for many key reactions. In addition to chlorine isomerizations, =CCI–CI dissociations, H abstraction by CI atoms, Cl atom transfer reactions, and radical addition reactions all represent major classes of reactions for which a lack of accurate rate data exists. Although serious attempts have been made to estimate the rates of these reactions, a larger database of validated CHC reaction rate data are needed to improve the development of extended reaction mechanisms.

The current status of PCDD/F formation modeling is not comprehensive and requires major improvement. For the current high-temperature PCDD model [95], additional PCDD decomposition/oxidation reactions may significantly impact the PCDD yields. Direct condensation reactions of molecular chlorinated phenols can be important at high temperatures [101, 102]. For PCDD/F modeling in general, there is a critical need for high-level calculations of rate constants and polynomials for key reactions as well as new pathways [103] and precursors [104]. The large uncertainties of rate parameters for reactions involving aromatic species as well as calculating thermodynamic parameters of these species may be the most difficult challenge in further improving PCDD/F formation models.

Bibliography

Primary Literature

1. Dempsey CR, Oppelt ET (1993) Incineration of hazardous waste: a critical review update. *J Air Waste Manage Assoc* 43:25
2. Electronic Code of Federal Regulations (1984) Appendix VIII to Part 261—Hazardous Constituents. http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&tpl=/ecfrbrowse/Title40/40cfr261_main_02.tpl. Accessed 11 July 2011
3. US-EPA (1983) Guidance manual for hazardous waste incinerator permits. U.S. EPA Office of Solid Waste and Emergency Response, Washington, DC, pp 2.16–2.26, NTIS PB84-100577
4. Lee KC, Morgan N, Hansen JL, Whipple GM (1982) Revised model for the prediction of the time-temperature requirements for thermal destruction of dilute organic vapors and its usage for predicting compound incinerability. In: *Proceedings of the seventy-fifth air pollution control association meeting*, New Orleans, LA, Paper 82–5.3
5. Tsang T, Shaub W (1982) Chemical process in the incineration of hazardous materials. In: Exner J (ed) *Detoxification of hazardous waste*. Ann Arbor Press, Ann Arbor, pp 41
6. Dellinger B, Rubey WA, Hall D, Graham JL (1986) Incinerability of hazardous wastes. *Hazard Waste Hazard Mater* 3:139
7. Trenholm A, Gorman P, and Jungelaus G (1984) Performance Evaluation of Full-Scale Incineration, Volumes I–V, US-EPA Final Report, EPA-600/2-84-181a–e. Cincinnati, OH
8. Clark WD, Seeker WR, Lee CC (1988) Engineering analysis of hazardous waste incineration: energy and mass balance. In: Freeman HM (ed) *Incinerating hazardous wastes*. Technomic Publishing, Lancaster, p 357
9. Dellinger B, Torres J, Rubey WA, Hall D, Graham JL, Carnes R (1984) Determination of the thermal stability of selected hazardous organic compounds. *Hazard Waste Hazard Mater* 1:137
10. Graham JL, Hall D, Dellinger B (1986) Laboratory investigation of thermal degradation of a mixture of hazardous organic compounds – I. *Environ Sci Technol* 2:703
11. Taylor PH, Dellinger B (1988) Thermal degradation characteristics of chloromethane mixtures. *Environ Sci Technol* 22:438
12. Dellinger B, Graham M, Tirey DA (1986) Predicting emissions from the thermal processing of hazardous wastes. *Hazard Waste Hazard Mater* 3:293
13. Taylor PH, Dellinger B, Lee CC (1990) Development of a thermal stability based ranking of hazardous organic compound incinerability. *Environ Sci Technol* 24:316
14. Carroll GJ, Thurnau RC, Lee JW, Waterland LR, Dellinger B, Taylor PH (1992) Pilot-scale evaluation of an incinerability ranking system for hazardous organic compounds. *J Air Waste Manage Assoc* 42:1430
15. Dellinger B, Taylor PH, Lee CC (1993) Full-scale evaluation of the thermal stability-based hazardous organic waste incinerability ranking. *J Air Waste Manage Assoc* 43:203
16. Taylor PH, Chadbourne JF (1987) Sulfur hexafluoride as a surrogate for monitoring hazardous waste incinerator performance. *J Air Pollut Control Assoc* 37:729
17. Dellinger B, Tirey DA, Taylor PH, Pan J, Lee CC (1988) Products of incomplete combustion from the high temperature pyrolysis of chlorinated methanes. In: *3rd Chemical congress of North America and the 195th national meeting of the American Chemical Society*, vol 28. Division of Environmental Chemistry Preprints, Toronto, p. 81

18. Chang DPY, Sorbo NW, Law CK, Steeper RR, Richards MK, Huffman GL (1988) Relationships between laboratory and pilot scale combustion of some chlorinated hydrocarbons. In: Proceedings of AIChE national meeting, Denver, CO
19. Cundy VA, Lester TW, Sterling AM, Montestruc AN, Morse JS, Leger CB, Acharya S (1989) Rotary kiln incineration iii an in-depth study – kiln exit/afterburner/stack train and kiln exit pattern factor measurements during liquid CCl₄ processing. J Air Waste Manage Assoc 39:944
20. Dellinger B, Taylor PH, and Tirey DA (1991) Minimization and Control of Hazardous Combustion Byproducts, US-EPA Final Report, EPA/600/52-90/039. Cincinnati, OH
21. Russell JJ, Seetula JA, Gutman D, Senkan SM (1989) Kinetics of reactions of chlorinated vinyl radicals (CH₂CCl and C₂Cl₃) with molecular oxygen. J Phys Chem 93:1934
22. Schuetzle D, Perez JM (1983) Factors influencing the emissions of nitrated Polynuclear Aromatic Hydrocarbons (PNAs) from diesel engines. J Air Pollut Control Assoc 33:751
23. Dickson LC, Lenoir D, Hutzinger O (1992) Quantitative comparison of *de novo* and precursor formation of polychlorinated Dibenzo-p-Dioxins under Simulated municipal solid-waste incinerator postcombustion conditions. Environ Sci Technol 26:1822
24. Addink R, Olie K (1995) Mechanisms of formation and destruction of polychlorinated Dibenzo-p-dioxins and Diben-zofurans in heterogeneous systems. Environ Sci Technol 29:1425
25. Fiedler H (1998) Thermal formation of PCDD/PCDF – a survey. Environ Eng Sci 15:49
26. Taylor PH, Sidhu SS, Rubey WA, Dellinger B, Wehrmeier A, Lenoir D, Schramm K-W (1998) Evidence for a unified pathway of dioxin formation from aliphatic hydrocarbons. In: Proceedings of the Combustion Institute, vol 27, pp 1769. Pittsburgh, PA
27. Stanmore BR (2004) The formation of dioxin in combustion systems. Combust Flame 136:398
28. Konduri R, Altwicker E (1994) Analysis of timescales pertinent to Dioxin/Furan formation on fly ash surfaces in municipal solid waste incinerators. Chemosphere 28:23
29. Gullett BK, Lemieux PM, Dunn JE (1994) Emissions of polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzofurans from the open burning of household waste in barrels. Environ Sci Technol 28:107
30. Golodets GI (1983) Heterogeneous catalytic reactions involving molecular oxygen. Elsevier, New York Volume 15
31. Pieters WJM, Conner WC, Carlson J (1984) The oxyhydro-chlorination of methane on fumed silica-based Cu⁺, K, La catalysts: I. Catalyst synthesis. Appl Cat 11:35
32. Wehrmeier A, Lenoir D, Sidhu SS, Taylor PH, Rubey WA, Ketrup A, Dellinger B (1998) Role of copper species in chlorination and condensation reactions of acetylene. Environ Sci Technol 32:2741
33. Nganai S, Lomnicki S, Dellinger B (2009) Ferric oxide mediated formation of PCDD/Fs from 2-monochlorophenol. Environ Sci Technol 43:368
34. Nganai S, Lomnicki S, Dellinger B (2011) Formation of PCDD/Fs from the copper oxide-mediated pyrolysis and oxidation of 1,2-dichlorobenzene. Environ Sci Technol 45:1034
35. Taylor PH, Lenoir D (2001) Chloroaromatic formation in incineration processes. Sci Total Environ 269:1
36. Trenholm A, Lapp T, Scheil G, Coores J, Klamm S, Cassady C (1988) Total mass emissions from a hazardous waste incinerator. J Hazard Mater 18:99
37. Ramamoorth S, Ramamoorth S (1997) Chlorinated organic compounds in the environment: regulatory and monitoring assessment. Lewis, New York
38. Aubrey NE, van Wazer JR (1964) Equilibrium rearrangement of perchlorocarbon compounds. J Am Chem Soc 86:4380
39. Choudry GC, Hutzinger O (1983) Mechanistic aspects of thermal formation of halogenated organic compounds including polychlorinated dibenzo-*p*-dioxins. Gordon and Breach, New York
40. Tirey DA, Taylor PH, Dellinger B (1990) Products of incomplete combustion from the high temperature pyrolysis of chlorinated methanes. In: Clement R, Kagel R (eds) Emissions from combustion processes: origin, measurement and control. Lewis, Chelsea, p 109
41. Taylor PH, Dellinger B, Tirey DA (1991) Oxidative pyrolysis of CH₂Cl₂, CHCl₃, and CCl₄ – I: incineration implications. Int J Chem Kinet 23:1051
42. Taylor PH, Dellinger B (1991) Modeling of molecular growth chemistry from high-temperature chlorocarbon pyrolysis. Chem Phys Process Combust, The Combustion Institute, pp 58/1
43. Dellinger B, Taylor PH, Tirey DA (1989) Pathways of formation of chlorinated PICs from the thermal degradation of simple chlorinated hydrocarbons. J Hazard Mater 22:175
44. Taylor PH, Wehrmeier A, Sidhu SS, Lenoir D, Schramm K-W, Ketrup A (2000) Copper-catalyzed chlorination and condensation of acetylene and dichloroacetylene. Chemosphere 40:1297
45. Mulholland JA, Sarofim AF, Sosthikul P, Monchamp PA, Plummer EF, Lafleur AL (1992) Formation of perchloroaromatics during trichloroethylene pyrolysis. Combust Flame 89:103
46. Taylor PH, Tirey DA, Dellinger B (1993) High temperature studies of the pyrolysis of trichloroethene: formation of perchlorinated PAHs. In: Joint Eastern and Central States Meeting of the Combustion Institute, New Orleans, pp 312
47. Taylor PH, Tirey DA, Rubey WA, Dellinger B (1994) Detailed modeling of the pyrolysis of trichloroethene: formation of chlorinated aromatic species. Combust Sci Technol 101:75
48. Ballschmiter K, Kirschmer P, Zoller W (1986) Experiments in high temperature chemistry of organohalogenes. Chemosphere 15:1369
49. Tirey DA, Taylor PH, Kasner JH, Dellinger B (1990) Gas-phase formation of chlorinated aromatic compounds from the pyrolysis of tetrachloroethylene. Combust Sci Technol 74:137

50. Taylor PH, Tirey DA, Dellinger B (1996) A comprehensive kinetic model of the high temperature pyrolysis of tetrachloroethene. *Combust Flame* 104:260
51. Taylor PH, Tirey DA, Dellinger B (1996) The high temperature pyrolysis of hexachloropropene: kinetic analysis of pathways to formation of perchloro-arylbenzenes. *Combust Flame* 105:486
52. Taylor PH, Tirey DA, Dellinger B (1996) The high temperature pyrolysis of 1,3-hexachlorobutadiene. *Combust Flame* 106:1
53. Baillet C, Fadli A, Sawerysyn J-P (1996) Experimental study on the thermal oxidation of 1,3-hexachlorobutadiene at 500–1100°C. *Chemosphere* 32:1261
54. Cole JA, Bittner JD, Howard JB, Longwell JP (1984) Formation mechanisms of aromatic compounds in aliphatic flames. *Combust Flame* 56:51
55. Westmoreland PR, Dean AM, Howard JB, Longwell JP (1989) Forming benzene in flames by chemically activated isomerization. *J Phys Chem* 93:8171
56. Weissman M, Benson SW (1984) Pyrolysis of methyl chloride, a pathway in the chlorine-catalyzed polymerization of methane. *Int J Chem Kinet* 16:307
57. Dellinger B, Taylor PH (1990) The role of chlorine-induced inductive destabilization versus bond energy reduction on high temperature displacement reactions of chlorinated hydrocarbons. *Chem Phys Process Combust, The Combustion Institute*, pp 9/1
58. Liem AKD, van Zorge JA (1995) Dioxins and related compounds: status and regulatory aspects. *Environ Sci Pollut Res* 2:46
59. Thomas VM, Spiro TG (1995) An estimation of dioxin emissions in the United States. *Toxicol Environ Chem* 50:1
60. US-EPA Dioxin Reassessment, National Academy of Sciences 2004 draft. <http://www.epa.gov/ncea/pdfs/dioxin/nas-review/>. Accessed 11 July 2011
61. Shaub WM, Tsang W (1983) Dioxin formation in incinerators. *Environ Sci Technol* 17:721
62. Kilgroe JD, Lanier WS, von Alten TR (1991) Montgomery county incinerator test program: formation, emission, and control of organic pollutants. In: 2nd International conference on municipal waste combustion, Tampa
63. Froese KL, Hutzinger O (1996) Polychlorinated benzene, phenol, dibenzo-p-dioxin, and dibenzofuran in heterogeneous combustion reactions of acetylene. *Environ Sci Technol* 30:1009
64. Milligan MS, Altwicker ER (1996) Chlorophenol reactions on fly ash. Adsorption – desorption equilibria and conversion to polychlorinated dibenzo-p-dioxins. *Environ Sci Technol* 30:225
65. Altwicker EH, Schonberg JS, Konduri RKNV, Milligan MS (1990) Polychlorinated dioxin and furan formation in incinerators. *Hazard Waste Hazard Mater* 7:73
66. Ismo H, Kari T, Juhani R (1997) Formation of aromatic chlorinated compounds catalyzed by copper and iron. *Chemosphere* 34:2649
67. Olie K, Addink R, Schoonenboom M (1998) Metals as catalysts during the formation and decomposition of chlorinated dioxins and furans in incineration processes. *J Air Waste Manage Assoc* 48:101
68. Born JGP, Mulder P, Louw R (1993) Fly-ash mediated reactions of phenol and monochlorophenols – oxychlorination, deep oxidation and condensation. *Environ Sci Technol* 27:1849
69. Ryu JY, Mulholland JA, Kim DH, Takeuchi M (2005) Homologue and isomer patterns of polychlorinated dibenzo-p-dioxins and dibenzofurans from phenol precursors: comparison with municipal waste incinerator data. *Environ Sci Technol* 39:4398
70. Hatanaka T, Imagawa T, Takeuchi A (2003) Effects of copper chloride on formation of polychlorinated dibenzo-p-dioxins in model waste incineration. *Chemosphere* 51:1041
71. Altwicker ER (1991) Some laboratory experimental designs for obtaining dynamic property data on dioxins. *Sci Total Environ* 104:47
72. Milligan MS, Altwicker ER (1993) The relationship between *de novo* synthesis of polychlorinated dibenzo-p-dioxins and Dibenzofurans and low-temperature carbon gasification in fly ash. *Environ Sci Technol* 27:1595
73. Stieglitz L, Zwick G, Beck J, Roth W, Vogg H (1989) On the *de novo* synthesis of PCDD/PCDF on fly ash of municipal waste incinerators. *Chemosphere* 18:1219
74. Stieglitz L, Zwick G, Beck J, Bautz H, Roth W (1990) The role of particulate carbon in the *de novo* synthesis of polychlorinated dibenzodioxins and dibenzofurans in fly ash. *Chemosphere* 20:1953
75. Oeberg T, Bergbaeck B, Oeberg E (2007) Different catalytic effects by copper and chromium on the formation and degradation of chlorinated aromatic compounds in fly ash. *Environ Sci Technol* 41:3741
76. Karasek FW, Dickson LC (1987) Model studies of polychlorinated dibenzo-para-dioxin formation during municipal refuse incineration. *Science* 237:754
77. Cains PW, McCausland LJ, Fernandes AH, Dyke P (1997) Polychlorinated dibenzo-p-dioxins and dibenzofurans formation in incineration: effects of fly ash and carbon source. *Environ Sci Technol* 31:776
78. Ballschmiter K, Braunmiller I, Niemczyk R, Swerev M (1988) Reaction pathways for the formation of polychloro-dibenzodioxins (PCDD) and -dibenzofurans (PCDF) in combustion processes: II. Chlorobenzenes and chlorophenols as precursors in the formation of PCDD/PCDF in flame chemistry. *Chemosphere* 17:995
79. Alderman SL, Farquar GH, Poliakoff ED, Dellinger B (2005) An infrared and X-ray spectroscopic study of the reactions of 2-chlorophenol, 1,2-dichlorobenzene and chlorobenzene with model CuO/Silica fly ash surfaces. *Environ Sci Technol* 39:7396
80. Pandelova M, Lenoir D, Schramm K-W (2006) Correlation between PCDD/F, PCB, and PCBz in coal/waste combustion: Influence of various inhibitors. *Chemosphere* 62:1196
81. Sommeling PM, Mulder P, Louw R (1994) Formation of PCDFs during chlorination and oxidation of chlorobenzenes in chlorine/oxygen mixtures around 340°C. *Chemosphere* 29:2015

82. Addink R, Cnubben PAJP, Olie K (1995) Formation of polychlorinated dibenzo-p-dioxins dibenzofurans on fly-ash from precursors and carbon model compounds. *Carbon* 33:1463
83. Froese KL, Hutzinger O (1996) Polychlorinated benzene and polychlorinated phenol in heterogeneous combustion reactions of ethylene and ethane. *Environ Sci Technol* 30:998
84. Lenoir D, Wehrmeier A, Sidhu SS, Taylor PH (2001) Formation and inhibition of chloroaromatic micropollutants formed in incineration processes. *Chemosphere* 43:107
85. Ritter E, Bozzelli JW (1990) Chlorobenzene and dichlorobenzene reactions in hydrogen and in hydrogen oxygen mixtures. *Hazard Waste Hazard Mater* 7:103
86. Fadli A, Briois C, Baillet C, Sawerysyn J-P (1999) Experimental study on the thermal oxidation of chlorobenzene at 575–825°C. *Chemosphere* 38:2835
87. Shaub WM, Tsang W (1985) Overview of dioxin formation in gas and solid phases under municipal incinerator conditions. In: Keith LH, Rappe C, Choudhary G (eds) *Chlorinated dioxins and Dibenzofurans in the total environment*. II. Elsevier, St. Louis, p 469
88. Sidhu SS, Maqsd L, Dellinger B, Mascolo G (1995) The homogeneous gas-phase formation of chlorinated and brominated dibenzo-p-dioxin from 2,4,6-trichloro- and 2,4,6-tribromophenols. *Combust Flame* 100:11
89. Grotheer H-H, Louw R (1996) The reaction of phenoxy radicals with monochlorobenzene and its meaning for gas-phase dioxin formation in incineration. In: *Proceedings of the Combustion Institute*, Pittsburgh, PA, 26, pp 2405
90. Khachatryan LA, Burcat A, Dellinger B (2003) An elementary reaction kinetic model for the gas-phase formation of 1,3,6,8- and 1,3,7,9-tetrachlorinated dibenzo-p-dioxins from 2,4,6-trichlorophenol. *Combust Flame* 132:406
91. Babushok Y, Tsang W (2003) Gas-phase mechanism for dioxin formation. *Chemosphere* 51:1023
92. Khachatryan L, Asatryan R, Dellinger B (2003) Development of expanded and core kinetic models for the gas phase formation of dioxins from chlorinated phenols. *Chemosphere* 52:695
93. Khachatryan L, Asatryan R, Dellinger B (2004) An elementary reaction kinetic model of the gas phase formation of polychlorinated dibenzofurans from chlorinated phenols. *J Phys Chem A* 108:9567
94. Procaccini C, Bozzelli JW, Longwell JP, Smith KA, Sarofim AF (2003) Formation of chlorinated aromatics by reactions of Cl, Cl₂, and HCl with benzene in the cool-down zone of a combustor. *Environ Sci Technol* 37:1684
95. Khachatryan L, Burcat A, Dellinger B (2010) The role of chlorine atoms and hydroxyl radicals in the formation of PCDDs from the oxidative pyrolysis of 2,4,6-trichlorophenol. *Int J Chem Kinet* 42:90
96. Babushok V, Tsang W, Noto T (2000) Propargyl-type radicals as precursors for polychlorinated aromatic hydrocarbons during incineration. *Proc Combust Inst*, Pittsburgh, PA, 28:2691
97. Sidhu SS, Dellinger B (1997) The effect of hydrocarbons on PCDD/F formation in the gas-phase oxidation of 2,4,6-trichlorophenol. *Organohalogen Compd* 31:469
98. Evans CE, Dellinger B (2003) Mechanisms of dioxin formation from the high-temperature pyrolysis of 2-chlorophenol. *Environ Sci Technol* 37:1325
99. Shinkura T, Kishi T, Kawakami T, Onodera S (2007) *Organohalogen Compd* 69:1602
100. Striebig RC, Rubey WA, Klosterman JR (2002) Trace-level measurement of complex combustion effluents and residues using multi-dimensional gas chromatography-mass spectrometry (MDGC-MS). *Waste Manag* 22:413
101. Okamoto Y, Tomonari M (1999) Formation pathways from 2,4,5-trichlorophenol (TCP) to polychlorinated dibenzo-p-dioxins (PCDDs): an ab initio study. *J Phys Chem A* 103:7686
102. Mohammednoor A, Bogdan Z, Dlugogorski Z, Kennedy EM, Mackie JC (2007) Theoretical study of reaction pathways of dibenzofuran and dibenzo-p-dioxin under reducing conditions. *J Phys Chem A* 111:7133
103. Okamoto Y (1999) A new dioxin decomposition process based on a hybrid density functional calculation. *Chem Phys Lett* 310:355
104. Farajian AA, Mikami M, Ordejon P, Tanabe K (2001) Ring closure in dioxin formation process: an *ab initio* molecular dynamics study. *J Chem Phys* 115:6401

Books and Reviews

- Brunner CR (1993) *Hazardous waste incineration*, 2nd edn. McGraw-Hill, New York. ISBN-10: 0-0700-8595-1
- Niessen WR (1995) *Combustion and incineration processes*, 2nd edn. Marcel-Dekker, New York. ISBN: 0-8247-9267-X
- Roberts SM, Teaf CM, Bean JA (1998) *Hazardous waste incineration: evaluating the human health and environmental risks*. CRC Press, Boca Raton. ISBN: 1-5667-0250-X
- Santolero JJ, Reynolds J, Theodore L (2000) *Introduction to hazardous waste incineration*, 2nd edn. Wiley, New York. ISBN-10: 0-4710-1790-6
- Waste Incineration and Public Health (2000) *Committee on health effects of waste incineration*, Board on Environmental Studies and Technology, National Research Council. The National Academies, Washington, DC. ISBN-10: 0-309-06371-X

Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds

LARS HÅKANSON

Department of Earth Sciences, Uppsala University, Uppsala, Sweden

Article Outline

Glossary

Definition of the Subject

Introduction, Background, and Aim
 A Brief Overview
 Key Catchment Characteristics
 Remedial Measures for Aquatic Pathways
 Future Directions
 Bibliography

Glossary

Adsorption The process of attachment on to and release from surfaces.

Algae Simple photosynthetic nonvascular plants; mainly aquatic.

Alkalinity The total amount of weak acid salts (largely bicarbonate) per unit volume of water.

Allochthonous From outside (often used to characterize materials transported from the drainage area to a lake).

Anoxic sediment Sediment devoid of free oxygen.

Autochthonous Originating from inside the lake.

Benthos Organisms living in or on the river or lake bed.

Bioaccumulation The increase in the amount of chemical or compound in the tissues of living organisms.

Biological half-life The time required for the amount of a particular radionuclide in a biological system, such as an animal, to be reduced by one half by biological processes, when the contamination has been terminated.

Biomass Mass of living organisms present at any one time within a given area or volume.

Biota The total flora and fauna of a given area.

Bioturbation Mechanical mixing, e.g., of bottom sediments due to living organisms.

Catchment The drainage basin that channels precipitation into a lake or single outflow.

Compartment Any part of the environment or process that may conveniently be considered as a single entity. Used in developing mathematical models.

Compartment model A model in which a series of compartments is used to represent the system of interest. Material can flow between the compartments. Differential or difference equations often are used to represent the rates of flow in the system.

Conductivity Quantity of electricity transferred across unit area per unit of potential gradient and unit of time (in mS/m). It usually gives an indication of total ionic concentration.

Confidence interval An interval that encompasses the true value for a parameter or measurement with a degree of confidence stated in terms of a probability.

Default value A value prescribed for a model parameter in the absence of data directly relevant to the assessment situation.

Diffusion The movement of atoms or molecules from a region of higher concentration of the diffusing species to regions of lower concentration.

Dimictic A lake having two seasonal overturn periods of mixing and two of thermal stratification (see thermocline).

Dynamic model A model that simulates the changes that occur through time in a system, especially, in the current context, in containment concentrations in the system.

Dystrophic Lake rich in humic matter mainly in the form of suspended plant colloids and larger plant fragments.

Ecological half-life The time required for the amount of a particular radionuclide in a particular organism living in a natural ecosystem to be reduced by one half.

Ecosystem A community of organisms together with the environment they inhabit and with which they interact.

Effective half-life The time required for the amount of a particular radionuclide in a system to be reduced by one half as a consequence of radioactive decay and all other processes.

Epilimnion The upper layer of lake water above the thermocline with a comparatively homogeneous temperature profile.

Eutrophic A lake rich in nutrients, usually resulting in high productivity.

Evapotranspiration Loss of water by evaporation from soil and by transpiration from vegetation over a given area with time.

Food chain Sequence of organisms in an ecosystem occupying specific hierarchical levels (trophic levels) such that organisms belonging to a superior level survive by eating organisms

belonging to inferior levels. The sequence can be represented as compartments in a mathematical model or analysis.

Food web The network of interconnected food chains in an ecosystem.

Generic model A model that is not built on site-specific information.

Hardness (water) A measure of the amount of calcium and magnesium cations in the water.

Hydraulic residence time The ratio of the volume of a water body to water discharge (corresponding to the theoretical time needed for a complete exchange of water).

Hypolimnion The layer of lake waters below the thermocline.

Load (internal/primary) Fallout directly onto the lake surface.

Load (secondary) The flux (transport) from the catchment to the lake.

Macrophytes Large aquatic plants.

Mesohumic Lake with intermediate levels of humic matter.

Monomitic Lake having a single period of free circulation or overturn per year.

Monte Carlo A technique involving the use of random numbers in a computer program or simulation model to represent stochastic events.

Morphometric All the geometric characteristics relative to the shape of the basin (volume, surface, depth, etc.).

Oligohumic A lake poor in nutrients, usually resulting in low productivity.

Oligotrophic A lake with low primary productivity.

Output variables Variables that are produced as model output. Output variables can be state variables or functions of one or more state variables.

Partition coefficient (K_d) The ratio between the amount of radionuclide attached to suspended matter per unit of mass and the amount of dissolved radionuclide per unit of volume of water.

Permeability Rate of passage of water through a given cylindrical section of a sediment core.

Pelagic Pertaining to the water column of the lake; used for organisms inhabiting the open waters of the lake.

Periphyton A community of plants, animals, and associated detritus forming a surface coating on stones, plants, and other submerged objects.

Phytoplankton Plant members of the plankton.

Piscivorous Fish eating.

Plankton Organisms that float or swim very feebly in the water masses of lakes and rivers.

Predator An animal that kills other animals for food.

Primary productivity The total production by photosynthetic and chemosynthetic activity of organic substances.

Resuspension The remobilization of particles from the sediments of a water body by the action of water movement.

Sediment The matter that has fallen to the bottom of a water body.

Sedimentation rate The amount of matter deposited on the bottom of a water body per unit of surface and time.

Sensitivity analysis The analysis of variation of model output with changes in the values of model parameters.

Stratification Division of lake water masses into horizontal layers with different physical and chemical properties.

Suspended matter Matter suspended in lake and river waters.

Thermocline The boundary layer of lake waters in which temperature changes sharply with depth; situated between the epilimnion and hypolimnion.

Tributary A smaller river flowing into a larger river or a lake.

Trophic level The hierarchical level that an organism occupies in the food chain. The group of organisms that occupy the same level in the food chain.

Uncertainty analysis An analysis of the way in which the uncertainty in assessment results is affected by uncertainty in the input data used in the model.

Validation (model) The process of comparing model outputs with independent experimental data sets. A model is considered validated when sufficient testing has been performed to ensure an acceptable level of predictive accuracy over the range of conditions over which the model may be applied.

Volume development The ratio between the lake volume and the volume of a cone having the basis equal to the area of lake surface and the height equal to the maximum depth of the lake.

Zooplankton Animal members of the plankton.

Definition of the Subject

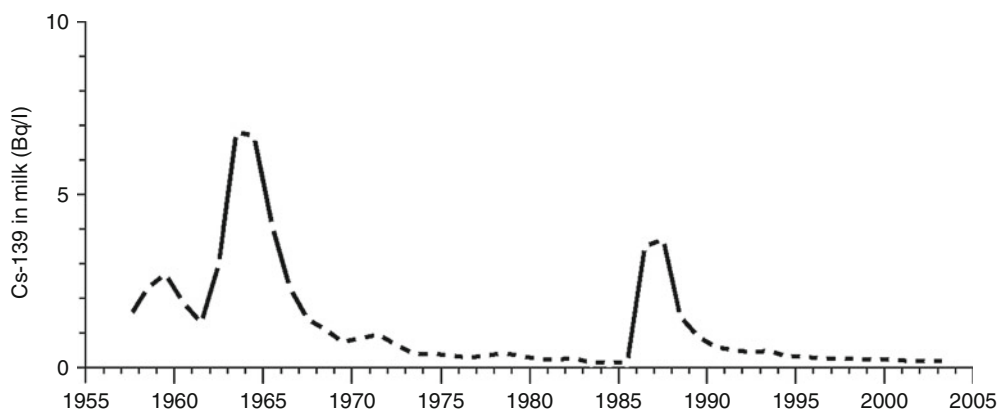
Aquatic radioecology deals with the transport processes of radionuclides in lakes, rivers, and marine areas. During the Chernobyl accident in April/May 1986, large amounts of radionuclides were released as a single pulse. To follow that pulse through ecosystem pathways has meant that many of the complex transportation routes can now be calculated with a much higher level of certainty than before.

Introduction, Background, and Aim

Figure 16 in Ed Waller's contribution "Sources of Radiation in the Environment" shows the fallout map of radiocesium (Cs-137) in soil systems throughout Europe. Note that there is no information from certain countries shown in white, e.g., for several Balkan states. Scandinavia (Norway, Sweden, and Finland) received a relatively large fallout with maximum values around 100 kBq/m² in the eastern part of central Sweden. The fallout was very patchy due to the prevailing atmospheric conditions, which were responsible for transporting the two main radionuclides, radiocesium and radiostrontium, from the site of the accident (Chernobyl) to basically all European countries. Figure 16 (Waller – Introduction) gives a spatial perspective to the problem and Fig. 1 a temporal perspective to the Chernobyl accident using data from Scandinavia. This figure exemplifies radiocesium

concentrations in cow milk in Sweden from the late 1950s to 2005. One can note the high and broad initial peak with maximum values in 1964 related to the nuclear weapons testing. It should be stressed that there were no such tests in Sweden or Scandinavia and this figure illustrates that radionuclides from these tests were distributed all around the globe. There is also a very pronounced peak from the Chernobyl fallout in April/May 1986. From these two background figures, the following questions will be addressed in this entry:

- "Peak and tail." A fundamental concern in radioecology is to understand, model, and predict the "peak and the tail" concerning radionuclide concentrations in ecosystems related to a given accidental fallout [1]. Peak concentrations are related to short-term conditions (hours to days after the fallout), whereas the "tail" concentrations generally are governed by the long-term conditions in the catchment areas that regulate fixation, percolation, surface and ground water transport, resuspension, and biotic fluxes of radionuclides [1–3].
- The literature related to the transport of radionuclides from the Chernobyl accident from land to water, in rivers, in lakes, and to coastal marine systems is extensive. Excellent compilations have been presented, e.g., for Scandinavia by [4], for the conditions in the Ukraine by [5]; and comprehensive review papers related to countermeasures



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 1

Radiocesium concentrations in Swedish milk between 1955 and 2005. Values in Bq/l (Modified from [98])

by [6], lake processes by [7], catchments by [8], and rivers by [9]. This entry will discuss key factors regulating the fluxes of radiocesium in watersheds and it will draw much from those reviews, but this article does not give a comprehensive literature review. The focus is on ecosystem function.

It should also be noted that during the last 10–15 years, there has been something of a “revolution” in ecosystem modelling. The generality and predictive power of present-day models have increased in a way that was inconceivable 15 years ago. A new generation of predictive models that predict as well as one can measure – if one measures well – have been presented. And yet, they may be run by readily available driving variables and have a general structure that applies to most types of pollutants in aquatic systems. The major reason for this development is, in fact, the Chernobyl accident. Large quantities of mainly radiocesium and radiostrontium were released in April/May 1986 as a pulse. To follow the pulse of these radionuclides through ecosystem pathways has meant that important fluxes and mechanisms, i.e., ecosystem structures have been revealed. It is important to stress that many of these new structures and equations are valid not just for radionuclides, but for most types of contaminants, e.g., for metals, nutrients, and organic radionuclides. This means that many of the models and methods of building and testing models discussed in this entry for aquatic systems should also be of interest to other ecosystem modellers.

Predictive models are fundamental tools for forecasting the consequences of different remedial actions (see [1]). The term “predictive model” is used to indicate models in which one or a few important y-variables are predicted from a few x-variables that can be obtained easily from standard maps or monitoring programs. That is the goal, but that goal is often very difficult to achieve. It is as much a challenge to develop practical models as it is to create comprehensive validated ecosystem models that predict the time-dependent interactions of many y-variables from many x-variables.

It should be stressed that many of the processes regulating the flow and biological uptake of radionuclides in aquatic systems are controlled by hydrological and morphometrical factors, which cannot be

influenced by remedies that change the water chemistry, e.g., liming and potash treatment [10–12]. Remedies that could speed up the recovery (e.g., reduction of the substances in fish) must aim at reducing the radionuclide load to the lake, or the uptake in biota by blocking the biouptake, increasing bioproduction and cause a “biological dilution” of the radionuclides in biota, and/or include specific substances, like potash (K), that can be taken up in fish in almost the same way as radiocesium and cause a “chemical” dilution. Time is an important factor in such remediations. The sooner the remedy is launched the better the results generally gained [13]. Lakes suited to potash treatment should have low initial ionic K-concentrations [14].

Predictive models must satisfy some categorical features that make them reliable tools for environmental management:

- They must be characterized by a relevant and simple structure, i.e., involve the smallest possible number of driving variables.
- The values of the necessary driving variables should be easy to access and/or to measure.
- The models must be validated for a variety of circumstances showing a wide range of environmental characteristics.

In broad terms, the variables used in environmental models may be divided into two categories:

1. Variables for which site-specific data must be available, such as lake volume, mean depth, water discharge, and amount of suspended particulate matter in water.
2. Variables for which generic (= general) values are used because of the lack of site-specific data, e.g., the sedimentation rate and/or rates for internal loading (like diffusion and advection rates).

The variables belonging to the first category are often called “site-specific variables,” “environmental variables,” or “lake-specific variables.” They can generally be measured relatively easily and their experimental uncertainty should not significantly affect the overall uncertainty of the model predictions of the target variable(s). The second category, the “model variables” or “model constants,” are often difficult to access for each specific system, such as the transfer rates from the

sediment to the water, the deposition velocity of a given radionuclide from water to sediments, and the migration rate from catchment to lake. The model variables may contribute significantly to the model uncertainty unless they have been validated and reliable sub-models for their generic values established from critical tests.

Unfortunately, critically validated widely applicable models showing a high predictive power are difficult to develop. As a consequence, predictive models are generally [15] based on “collective parameters.” It has been shown that [16] in many circumstances, the values of such important collective parameters integrate many compensatory effects of the different phenomena occurring in the very complex ecosystem where “everything depends on everything else” [17]. Examples of such collective parameters in the freshwater environment are the “migration velocities” of the metal or radionuclide from water to sediments, the “effective removal” rates, and the “soil permeability coefficients” of a radionuclide from the catchment to a water body [18].

Models based on such “collective parameters” show a unique and important feature: Their predictions are characterized by a relatively low uncertainty, despite the large range of environmental characteristics and the lack of site-specific values of the model variables [19]. The use of models based on collective parameters is the result of a change in the perspective concerning the modelling of natural ecosystems. The main lesson is that in predictive modelling, it is seldom necessary, or wise, to account for “everything.” The difficult task is to omit processes that may add more to the model’s uncertainty than to the predictive success for the given target variable.

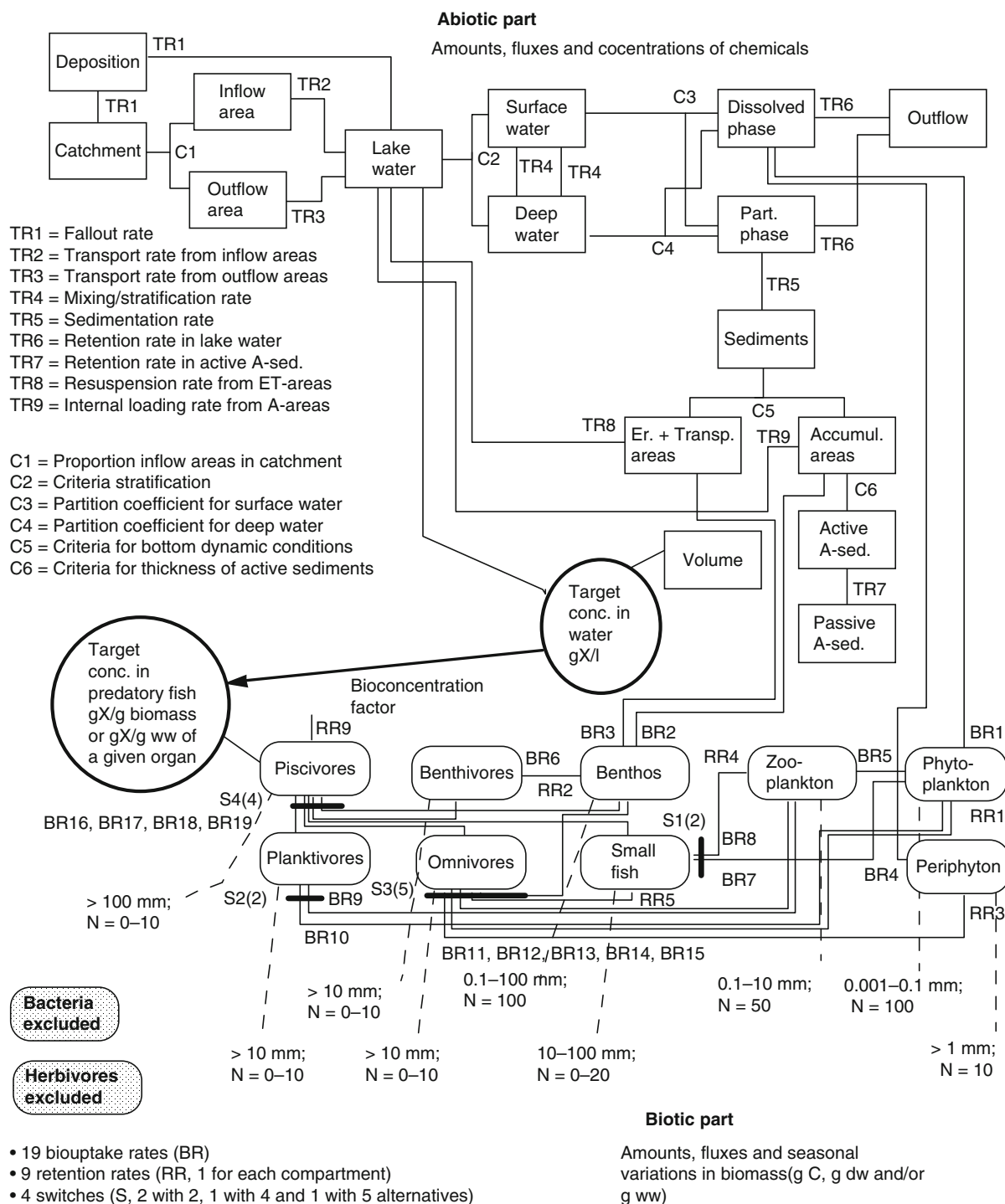
Environmental models based on collective parameters are structured more like a web than a pyramid. Each process is indeed related to a variety of other phenomena and it would be unreasonable to use only a few of them as fundamental starting points for understanding and predicting all the others. A variety of experiences [17, 20, 21] demonstrate that big complex models are often more uncertain, and yield poorer predictive success than simpler models based on collective parameters.

It is often extremely difficult to distinguish cause and effect in natural ecosystems. One cannot base

a dynamic ecosystem model on a full understanding of the ecosystem. In complex ecosystems, “understanding” at one scale (e.g., the ecosystem scale) is generally related to processes and mechanisms at the next lower scale (e.g., the scale of individual animals and/or plants), and the explanation of phenomena at this scale is related to processes and mechanisms at the next lower scale (e.g., the scale of the organ), and so on down to the level of the atom and beyond. In radioecology and environmental management, the predictive scientist must often find a balance between answering interesting, often important, questions of understanding, and delivering a practical tool to society. If a model were based on a causal analysis of what takes place at the cellular level, then at levels involving organs, individuals, populations, and finally at the ecosystem level, one would wait an eternity before the model could be developed.

In a strict sense, there is no such thing as a general (= generic) ecosystem model that works equally well for all ecosystems (of a given type) because all models need to be tested against reliable, independent empirical data and the data used in such validations must of necessity belong to a restricted domain. If this domain is equal to the entire population of ecosystems of the given type, then, and only then, is the model generic in the strict sense. The complexities of natural ecosystem always exceed the complexity and size of any model. Simplifications are always needed, and this entails problems. The dynamic mass-balance models discussed in this entry have been tested over such wide ranges that it is tempting to label them generic, but there will always be an ecosystem with properties outside the given domain for which the model would yield poor predictions. This is why modelling can be pictured as a two-sided coin: One needs the equations as well as the range, the domain, within which the equations apply.

The crucial point in ecosystem modelling is the structuring of the model see (Fig. 2 for an outline related to key concepts for lakes). The model structure should include all important processes regulating the transport and biouptake of a given radionuclide – no more, no less. The principles and processes illustrated in Fig. 2 will be further elaborated in the following text.



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 2

A compilation of fundamental abiotic and biotic processes regulating the spread and biouptake of toxic substances in aquatic ecosystems. The figure also gives typical sizes (mm) and numbers of species (N) in Nordic lakes. This structuring of the aquatic ecosystem will be used in the following model, where the concepts will be further elaborated (Modified from [63])

A Brief Overview

The aim of this section is to give a few empirical examples illustrating the temporal and special development for radiocesium and radiostrontium related to the Chernobyl accident. Figure 3 first shows measured mean levels of radiocesium in mouse shot in Heby commune in central Sweden (close to Uppsala and Stockholm) in the years 1986–2004. This figure illustrates some important principles related to the “peak and tail” problem. First, the recovery is relatively slow and principally governed by the physical half-life for radiocesium (30 years). So, the ecological half-life in this example is close to the physical half-life. Then, there are marked variations in the radiocesium levels in mouse from year to year depending on what the animals have been eating (“you are what you eat”). In years when mushrooms are abundant, the Cs-levels are generally high because many types of mushrooms with mycelium pick up radiocesium from the soils, and vice versa.

Figure 4 exemplifies river data from several European rivers (see Table 1) on annual concentrations of radiocesium from the Chernobyl fallout. One can note that among these river sites and years, the variation is

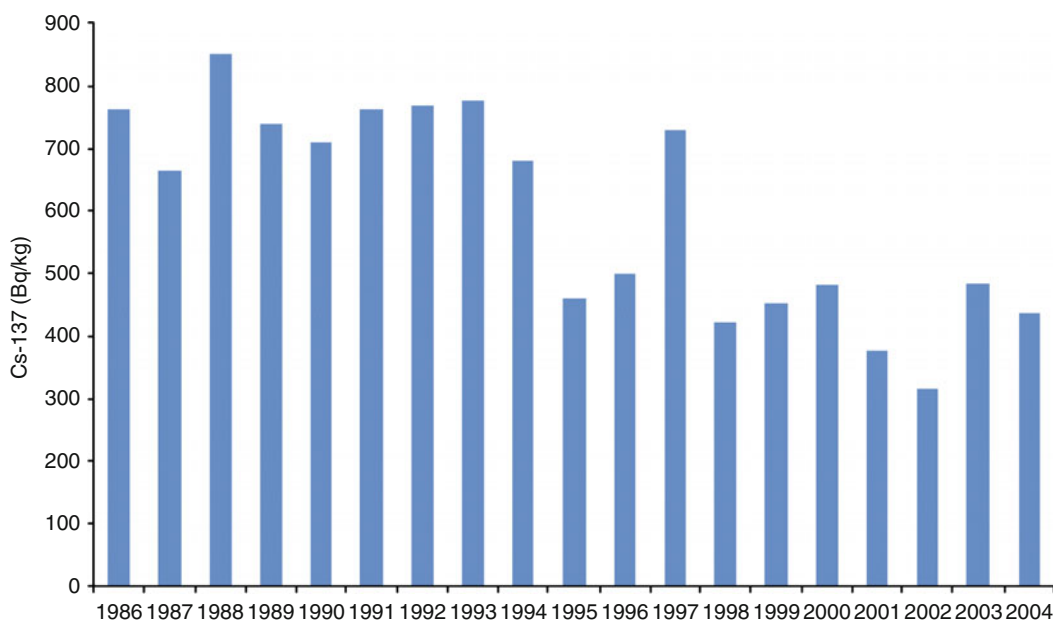
three to four orders of magnitude. The basic aim of the following discussion related to the transport from land to water is to highlight important factors regulating the variation among and within river sites. The first five rivers listed in Table 1 are Scandinavian (Finnish).

Figure 5 gives data on annual concentrations for radiostrontium from the nuclear weapon fallout and the Chernobyl fallout. The variability among the sites and the years is 100 times lower than for radiocesium.

Key Catchment Characteristics

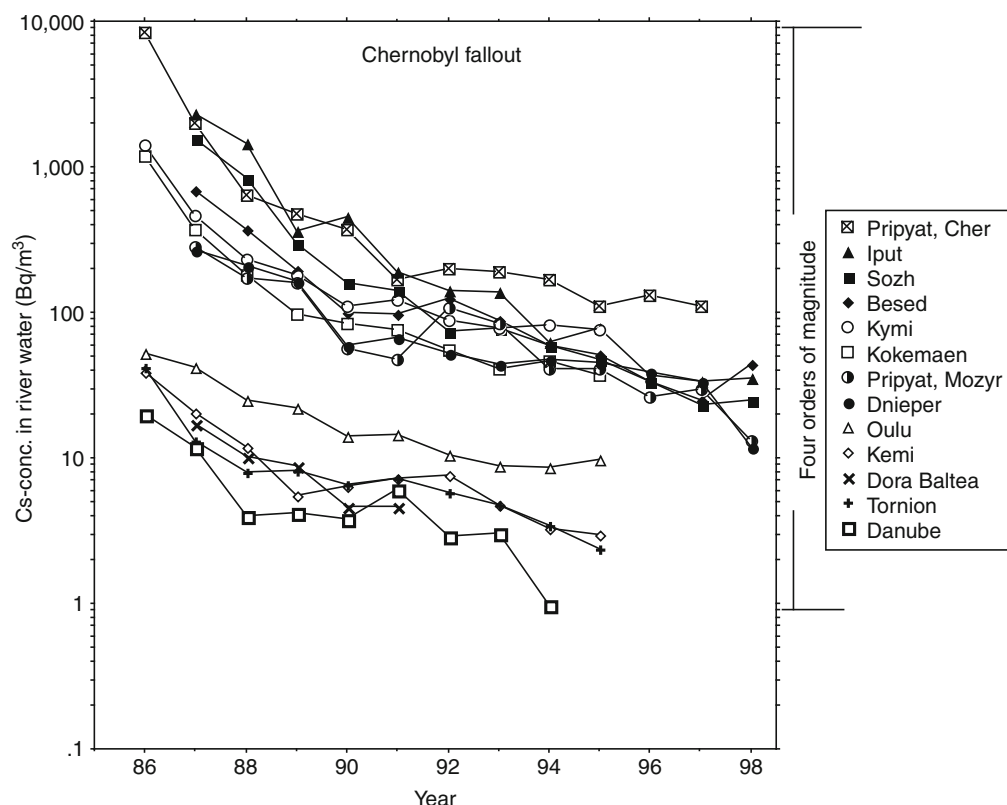
What are the key factors regulating the transport of radionuclides from land to water explaining the three to four orders of magnitude variability for radiocesium concentrations in rivers illustrated in Fig. 4? The first and most evident reason for the variability is the variability in fallout.

Many studies [22, 23] have demonstrated that aquatic systems are affected by catchment characteristics (e.g., area covered by bedrocks, soils, and different land-use activities). This is exemplified in Fig. 6 by river data on radiocesium. This figure gives normalized data (i.e., river water concentrations divided by fallout) and



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 3

Radiocesium concentrations in mouse from Heby commune (close to Uppsala in Sweden) between 1986 and 2004 (Modified from [99])



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 4

Measured radiocesium concentrations in rivers contaminated by the Chernobyl fallout. Note that these empirical data vary among the rivers sites and with time between 1 and 10,000 Bq/m³. The basic aim of the river model is to explain the factors regulating these variations (From [41])

one can see that the variations among the rivers sites is 10–100 times lower than in Fig. 4 without normalization for fallout. The remaining variation (a factor of 2) among the river sites depends on differences in catchment area and river characteristics.

There are basically two main factors related to the conditions in the catchment area, [1] the distribution of inflow and outflow areas and [2] fixation.

Figure 7 first illustrates principal aspects related to fixation for radiocesium and radiostrontium in catchment areas. From this figure, one can see that radiostrontium is more mobile, less fixed, than radiocesium. The fixation of any given radionuclide depends on:

1. The soil type.
2. The percentage of outflow areas (OA = wetlands like mires, bogs, upstream rivers, and lakes).

3. The mobility constant, which describes the mobility of a given radionuclide relative to the mobility of radiocesium. The default value for the dissolved fraction (DF) or radiostrontium in river water may often be set to 0.95 and for radiocesium to 0.56 see (Fig. 8). The higher the DF-value, the more mobile the substance, and this is also shown in Figs. 7 and 8.

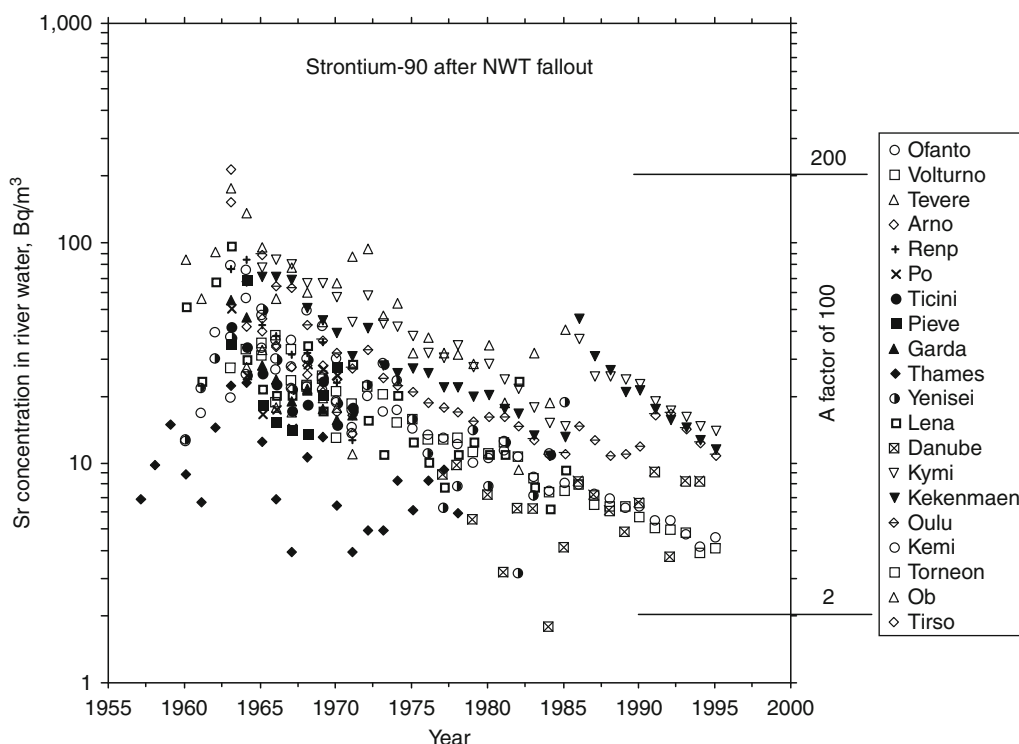
An important concept in the modelling of radionuclides concerns the equilibrium constant, the K_d -value between the amount of the radionuclide in dissolved and particulate phases [24–27]. The default value for lake K_d for radiocesium is often set to about 800,000 l/kg or 0.8 l/mg suspended solids (from [1]). Another approach for radiocesium (see [28–30]) is to express K_d as a function of the “frayed edge site” concentration of the clay mineral illite (which has a unique

Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Table 1 Data on radiocesium and radiostrontium fallout after Chernobyl and the nuclear weapons tests (NWT), catchment areas, mean annual precipitation (data from [78]), altitudes, and latitudes for the studied river sites. Data on latitude and altitude have been obtained from an atlas of Europe

River	Catchment area (km ²)	Annual precipitation (mm/year)	Chernobyl fallout (Bq Cs/m ²)	Total NWT deposition (Bq Cs/m ²)	Total NWT deposition (Bq Sr/m ²)	Altitude (m.a.s.l.)	Latitude (°N)
Kymi	33,620	566	29,900	3,500	2,180	0	63
Kokenmaen	26,620	549	34,100	3,400	2,110	0	62
Oulu	24,800	518	3,800	3,200	2,000	0	65
Kemi	42,760	467	1,700	3,060	1,800	0	66
Tornion	56,830	497	1,700	2,900	1,900	0	66
Thames	12,370	719			1,840	0	52
Tirso	3,360	867			3,340	0	44
Ofanto	3,390	847			3,260	0	41
Volturno	7,070	941			3,600	0	42
Tevere	17,290	1,038			4,000	0	42
Arno	11,020	1,119			4,300	0	43
Reno	7,820	1,163			4,478	0	43
Po (Torino)	9,190	987			3,800	240	45
Ticini	8,100	1,301			5,000	50	45
Dora Baltea	4,020	987	6,180			270	46
Pieve	4,890	1,731			6,660	0	45
Garda	13,900	1,380			5,300	100	46
Dnieper	50,610	696	38,900			200	52
Sozh	22,330	709	254,000			250	51
Besed	4,890	701	283,000			1,000	51
Iput	11,050	700	275,000			200	51
Pripyat, Mozyr	63,970	675	31,200			50	52
Pripyat, Che	46,020	694	138,000			100	52
Danube	180,810	962	11,100	5,920	3,700	300	47
Yenisei	2,564,580	487			1,880	0	70
Lena	2,397,580	418			1,600	0	72
Ob	2,950,760	477			1,840	0	67

potential to bind cesium), C_{FES} , the K-concentration of the water, C_K , and the selectivity coefficient between radiocesium and potassium (K_c), where $K_d = K_c \cdot C_{FES} / C_K \cdot K_c$ and C_{FES} are often assumed to be constants since it is generally very difficult to access reliable empirical

data for these variables for different lakes. This means that the approach may be written as $K_d = \text{const} / C_K$. On the basis of these arguments, the following may be regarded as a typical “collective” approach, i.e., as a simplification of many complex relationships [16],



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 5

Measured radiostrontium concentrations in rivers contaminated by the NWT (nuclear weapons testing) fallout and the Chernobyl fallout. These data vary among the rivers and with time between 2 and 200 Bq/m³. The aim of the river model discussed in the text is to account for the factors regulating these variations (From [41])

where positive “errors” are balanced by negative ones. Thus, for Cs-137:

$$K_d = 800,000/C_K \quad (1)$$

where C_K is given in mg K/l (K = potassium). 800,000 is the K_d -constant.

Traditionally, K_d is defined as is the ratio between the particulate (C_{part} in Bq/kg dw) and the dissolved (C_{diss} in Bq/l) phases, i.e., $K_d = C_{part}/C_{diss}$. The total amount (in Bq/l) is equal to $(C_{diss} + C_{part})$, where C_{part} is given in Bq/l. Then K_d is given in l/kg. The amount of suspended particulate matter in the lake water, SPM, is given in mg/l. This means that the dissolved fraction (DF) can be written as:

$$DF = 1/(1 + K_d \cdot SPM \cdot 10^{-6}) \quad (2)$$

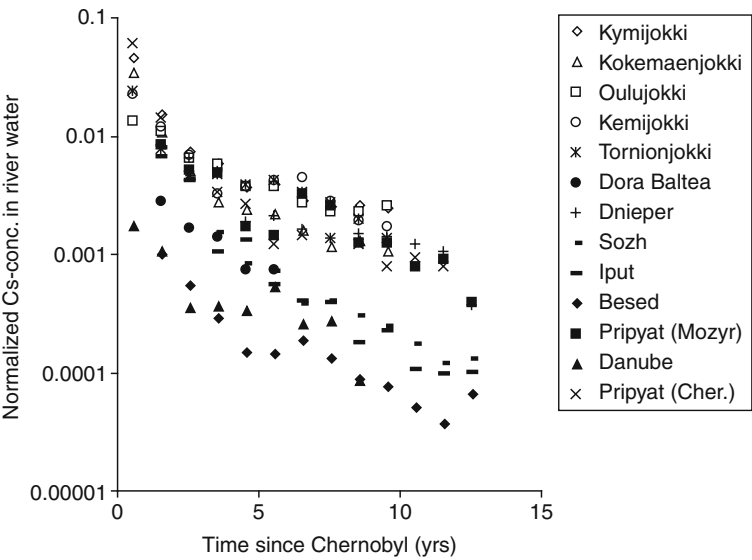
Since C_K generally varies in the range from 0.5 to 50 mg/l in aquatic systems [31] and since SPM-values generally also vary in the range between 0.5 and

50 mg/l [32], it is evident that DF varies in the entire range from almost 0–1. For example, for lakes with characteristic C_K — and SPM-values between 1 and 5 mg/l, one can expect DF-values for radiocesium in the range from 0.25 to 0.75. Håkanson et al. [33] have shown that DF for Sr-90 may be given by this approach but that Eq. 1 is replaced by

$$K_d = C_{CsSr} \cdot 800,000/C_{Ca} \quad (3)$$

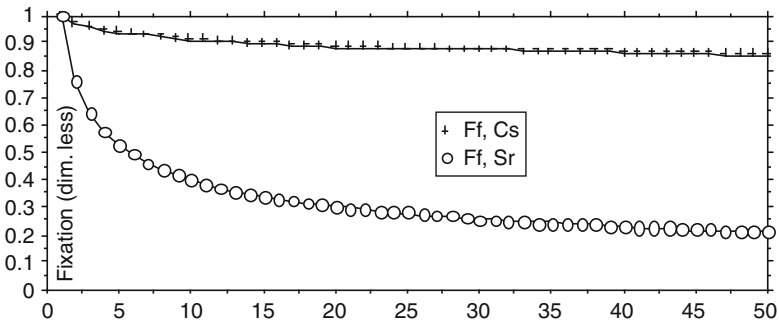
Where C_{CsSr} is a calculation constant between Cs and Sr set to 10 and C_{Ca} is the Ca-concentration in mg/l, since Ca should work in relation to Sr in the same competing manner as K does to Cs.

Figure 8 also illustrates that the principle of fixation/mobility should be valid not just for radionuclides but also for stable metals. The information of cation exchangeability illustrated for Hg, Pb, Cd, Cu, Zn, Cr, and Ni in Fig. 8 comes from the literature [32, 34, 35].



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 6

Total concentrations of Cs-137 in water in 13 European rivers after Chernobyl (mean annual values) normalized for fallout (Bq/m^3 to Bq/m^2). These normalized concentrations vary two orders of magnitude due to differences among the rivers sites in catchment area and river characteristics. The aim of the model discussed in the text is to quantify such differences based on mechanistic principles regulating the flow of radionuclides from land to water and the internal factors regulating the retention of radionuclides in rivers. The model gives monthly predictions of the target variable, radionuclide concentrations at defined sites in rivers (From [41])



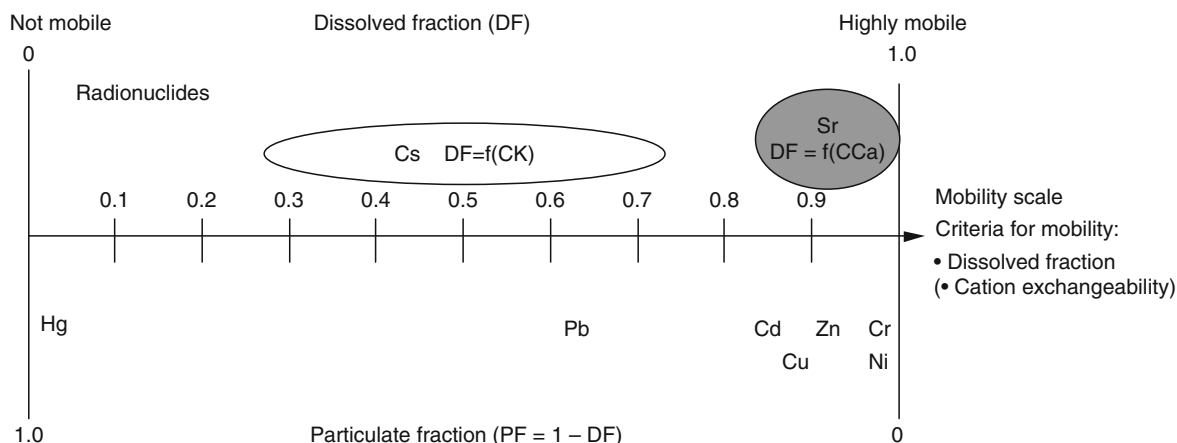
Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 7

Illustration of fixation in catchment areas calculated for radiocesium and radiostrontium (From [41])

The catchment area is often divided into two parts see (Fig. 9):

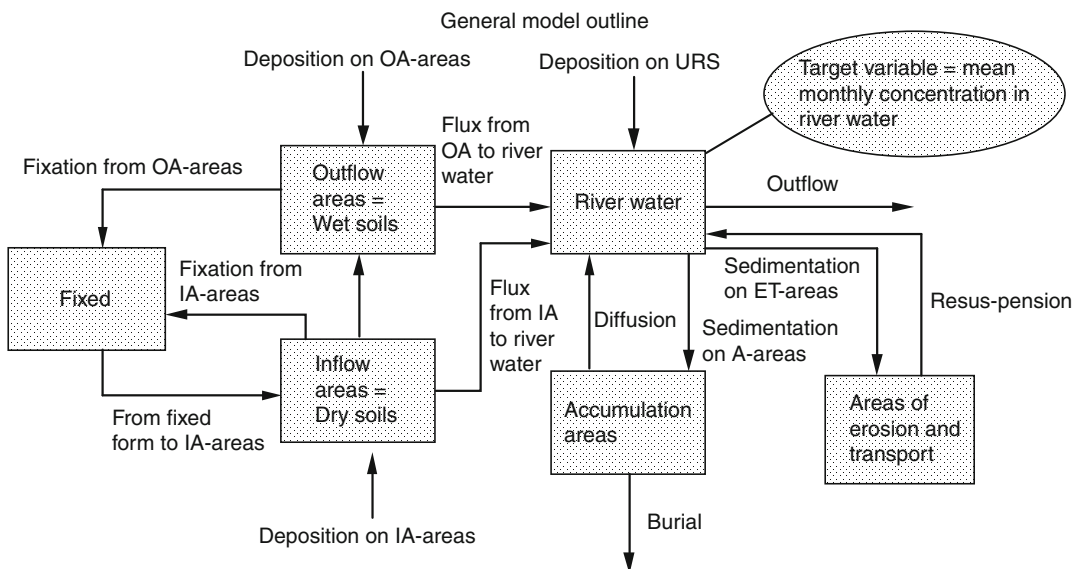
1. The outflow areas (i.e., wetlands = upstream lakes, rivers, bogs, and mires), where there may be a relatively quick horizontal transport of the radionuclide from land to surface water, and the remainder.
2. The inflow areas (= the dry land areas), where there is a much slower transport of the radionuclide first vertically to the ground water through several soil layers and then from the ground water to the surface water.

Many papers and textbooks in hydrology treat these topics thoroughly [22, 36–39].



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 8

A schematic illustration of the mobility and the dissolved fraction of various substances



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 9

An outline of the river model

Rivers Fluxes

The following processes are important for the river transport see (Fig. 9):

1. Direct fallout onto the river
2. Input from the catchment area (one flux from outflow areas to the river water, another flux from inflow areas)
3. Internal fluxes (sedimentation on ET-areas and A-areas; i.e., areas of fine sediment erosion plus transportation and fine sediment accumulation, respectively; see Section [Dynamic Modelling](#) for further explanation)
4. Resuspension of fine sediments and particulate forms of the radionuclides from ET-areas back to river water

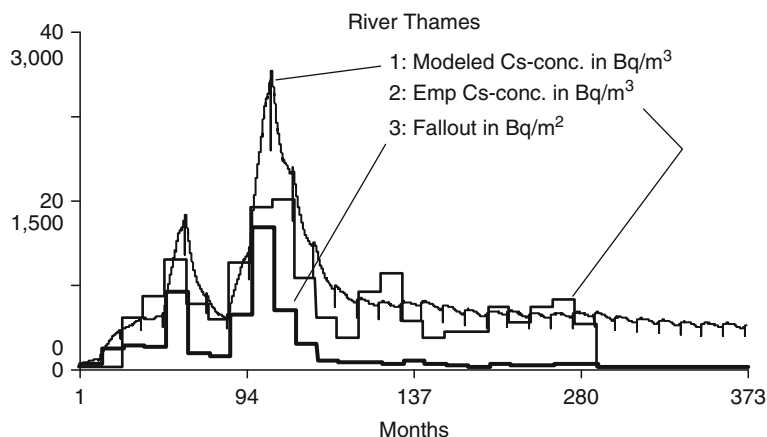
5. Diffusion from A-areas back to river water (i.e., the return flow from sediments back to water related to concentrations gradients; see [40]; this is especially important for radiocesium)
6. Outflow to downstream river stretches

Modelling Catchments and Rivers

There are uncertainties in most empirical data used to run and test models, e.g., in the climatological factors regulating precipitation and runoff of the radionuclides from land to water. This is exemplified in Fig. 10 using data for the River Thames. This figure gives the fallout (curve 3), the measured Cs-concentrations related to the fallout from the nuclear weapons test (NWT; curve 2), and the modelled Cs-concentrations in river water (curve 1; from [41]). Note that there is nothing in the fallout curves that would suggest that there should be a small peak in measured river concentrations just before week 280 (week 1 is January of 1955). This should be attributed to, e.g., hydrological/climatological factors. To be able to predict variations of this type, one would need actual data on precipitation and water discharge for the entire period. Such data are often not available, or very difficult to access for many/most rivers. Further, such data add more descriptive power rather than predictive power to the model, since one cannot predict relevant

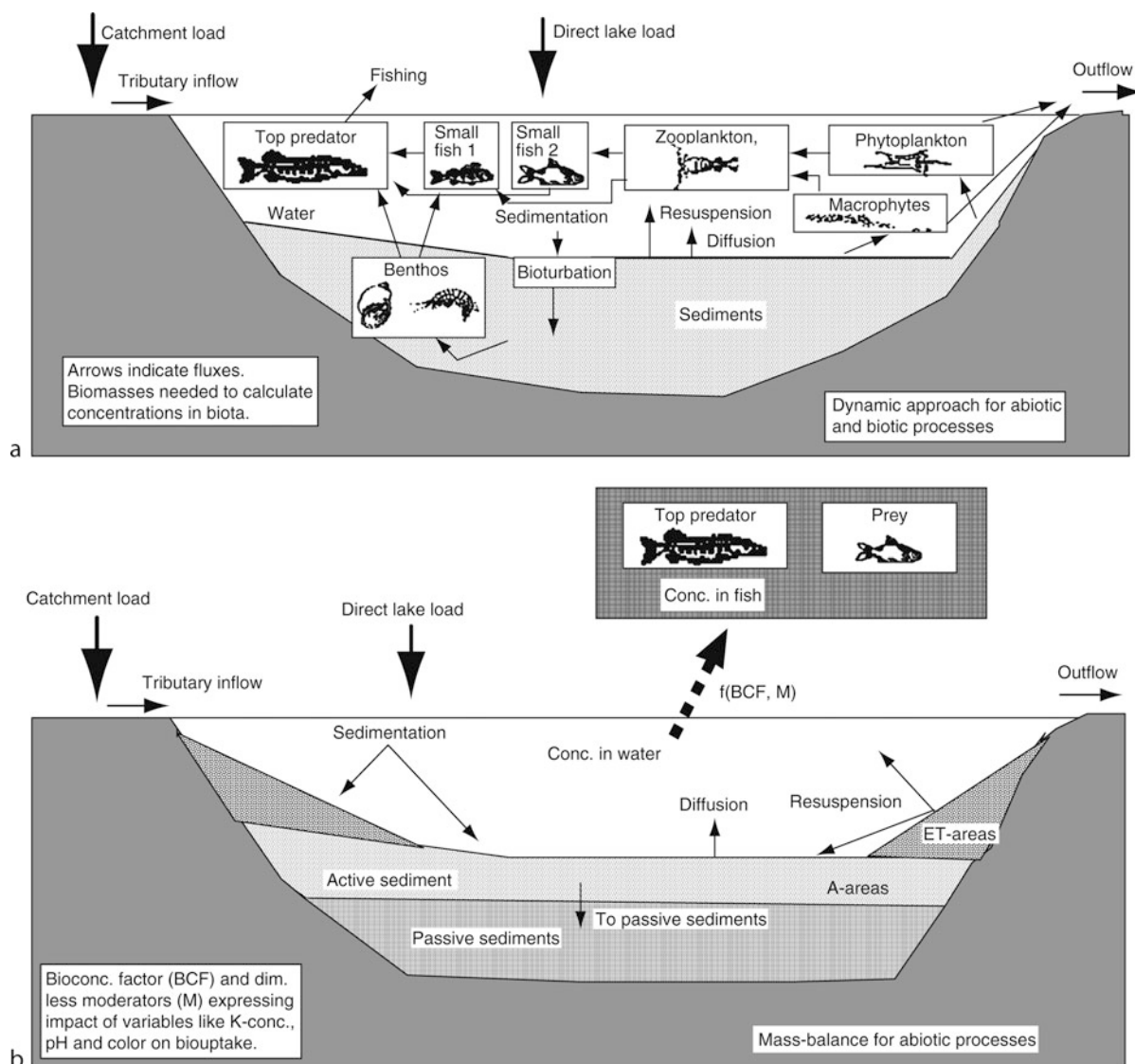
climatological conditions but more than a few days ahead. Given such uncertainties, one can understand that modelling of radiocesium would not predict concentrations in rivers better than a factor of 1.6 compared to the empirical data and not Sr-concentrations better than a factor of 2.2. The characteristic coefficients of variation ($CV = SD/MV$; SD = standard deviation, MV = mean value), as a measure of the uncertainties in the empirical data for radionuclide concentrations in rivers are generally simply too high and higher for radiostrontium than for radiocesium [41].

A graphical illustration of a traditional compartmental modelling approach for aquatic systems is given in Fig. 11a. This figure illustrates top predator (like pike or large perch), small fish 1 and 2 (e.g., small perch and roach), zooplankton, phytoplankton, algae and zoobenthos, and abiotic compartments (like water and sediments) and the processes regulating the fluxes between these compartments. The figure also gives the fluxes to the given aquatic system (direct lake load and river input related to catchment load) and from the lake (outflow and burial = sediment transport from active to passive sediment layers). This is a general modelling setup applicable for any substance X in a lake. The concentration of X in any biological compartment is defined by the ratio between the amount of



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 10

Data from River Thames illustrating the NWT fallout (curve 3), measured mean annual values of Cs-concentrations in river water (curve 2), and modelled Cs-concentrations (curve 1). Note that there are significant variations (a factor of 2) in the empirical data, which cannot be attributed to variations in fallout but must be explained by climatological variations in water discharge, runoff from land, etc. (From [41])



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 11

(a) Illustration of a traditional compartment model showing the fluxes (arrows; mass per unit time) for a given radionuclide or metal in a lake ecosystem with compartments for top predator, two types of small fish, zooplankton, phytoplankton, algae, benthos, water and sediments. (b) Illustration of another modelling approach showing a model based on mass-balance concepts for all abiotic fluxes, a bioconcentration factor, which is used to calculate concentrations of toxic substances in biota (so that one does not need to determine biomasses), and dimensionless moderators used to quantify how different water chemical conditions (pH, color, K-concentration, etc.) affect biouptake via the bioconcentration factor (From [19])

X in the compartment (in g or Bq) divided by the biomass of the species (e.g., kg pike in the lake). One needs biouptake rates and excretion/retention rates for all biological compartments, variables to describe the

compartments (like age, sex, organ, weight, and feed habits) and the biomasses of the species in the lake. All these variables are difficult and expensive to access; they vary seasonally and are often based on expert

opinion. This means that models of this kind are often based on many opinions and assumptions, which explain the fact that they generally provide poor predictions (see [17]). Better predictions can often be obtained for models structured in a different way. Then, one would use a dynamic (mass-balance) model for all important abiotic processes, which are relatively easy to handle compared to biological processes, so that the concentration of X in water and/or sediments can be predicted accurately, and then use the bioconcentration factor (BCF) approach to model the biological fluxes so that the concentration of X in biota is calculated directly without having to estimate biomasses. Often there is knowledge about the factors influencing the biouptake, like a lower biouptake for Cs-137 in lakes with high concentrations of potassium. This knowledge can be expressed by dimensionless moderators operating on the BCF-value see (Fig. 11b).

The inherent uncertainty for the Cs-concentrations in rivers, which is 1.6 [41], is illustrated in Fig. 12. Figure 12 gives a direct comparison between modelled concentrations of radiocesium against empirical data for the European rivers listed in Table 1. Most of the data, but not all, are within these two uncertainty bands for the empirical data. This means that the model (from [41]) can explain most of the variability in this situation (related to the Chernobyl fallout for Cs-concentrations in rivers; compare Fig. 4).

Modelling Lakes

Background Many studies have been done on especially radiocesium in lakes after the Chernobyl accident. It is evident that many factors and processes can influence the variability in radiocesium concentrations in water and biota within and among lakes. But what relative importance should one expect for the various factors and processes? It is evident that hundreds of x-variables may be of interest, but they cannot all be of equal importance for the predictions of the target y-variable (generally Cs-137 in fish eaten by man). Only reliable empirical data can be used to provide a quantitative ranking of different x-variables influencing a defined target y-variable. Table 2 may be regarded as an introduction to the following modelling. It gives (based on data from 14 Swedish lakes; see [3, 10, 11]) a ranking based on stepwise multiple

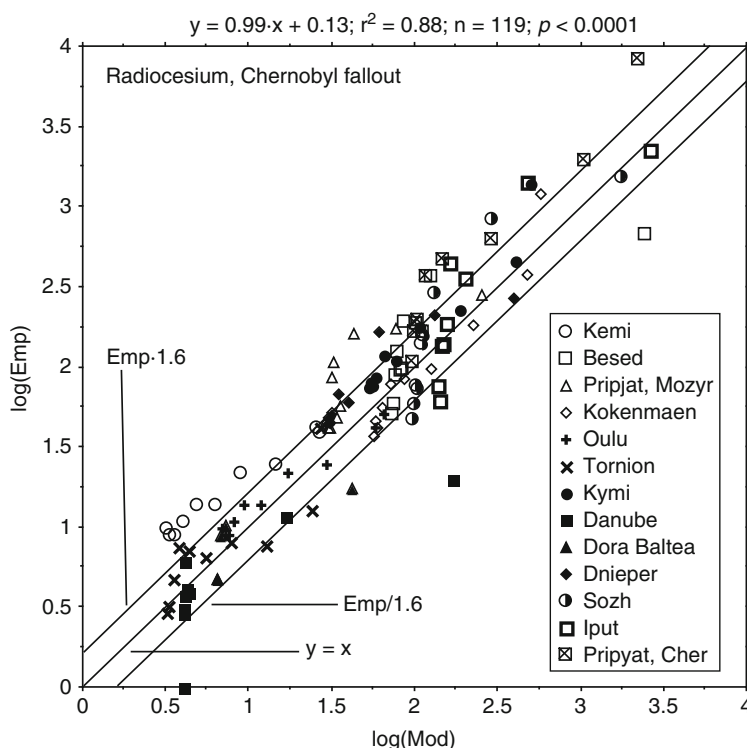
regression analysis of the factors influencing a target y-variable, here the concentration of radiocesium in pike (C_{pi88}) caught in the spring of 1988 in relation to:

1. One variable for the load of radiocesium to the lake, C_{wa87} (Cs-concentration in lake water in 1987 in Bq/l). Accounting for this explains statistically 78% of the variation among the lakes in C_{pi88} .
2. Different lake variables (mean annual values for 1987), first and logically, variations among the lakes in K-concentrations (in meq/l); the r^2 -value (the coefficient of determination) increases to 0.89 if this is accounted for.

It is evident that the target y-variable in this test, C_{pi88} , may also be related to the fallout (C_{soil}) after the Chernobyl accident, and several of the given water variables could, potentially, influence the bioavailability and biouptake of radiocesium as well as the biomasses, and hence the concentration of radiocesium in the biomasses. The morphometrical parameters could, potentially, influence the retention of radiocesium in lakes, the resuspension and the internal loading of Cs-137, and the catchment parameters could all, potentially, influence the runoff of radiocesium from land to water, i.e., the secondary load of radiocesium to the lakes. But all these factors cannot be of equal importance in predicting C_{pi88} .

It should be stressed that this empirical model only applies for a given year after the Chernobyl accident. The most important water chemical variable in this empirical model is the lake K-concentration – the more ions similar to Cs, like K, the lower the uptake of ^{137}Cs – a case of “chemical dilution” [42–44]. For a “peak emission” like this, the biouptake of Cs-137, and the Cs-concentration in fish, is also lower in lakes with fast water turnover than in lakes with slow water turnover – a case of normal “water dilution.” The higher the lake phosphorus concentration, the higher the bioproduction, the greater the biomasses, and the lower the Cs-concentrations in the biomasses – a case of “biological dilution.”

Radiocesium is rather reactive [25, 45–47]. This is very important and means that the behavior of radiocesium in lakes is strongly influenced by the chemical and physical properties of the carrier particles (illite clays). In addition, the retention in a lake will be



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 12

Results for radiocesium concentrations in river water related to the Chernobyl fallout. The figure also gives the ideal line ($y = x$), the two uncertainty lines related to the uncertainty in the data (measured river concentrations, fallout, and catchment features; note that these lines indicate that one cannot expect that the model should predict better than the interval given by these two lines). The slope of the regression line is 0.99, which is very close to the ideal and indicates that the model on average neither overpredicts nor underpredicts the empirical data. Over this wide range of data, the coefficient of determination (r^2) is 0.88. The number of measured data (n) is 119 and the statistical certainty (p) < 0.0001 (From [41])

Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Table 2 Results from stepwise multiple regression analyses (from [100]) based on data from 14 Swedish lakes. y-variable = radiocesium concentrations in pike caught in 1988 (C_{pi88} in Bq/kg ww) tested against different x-variables that could potentially influence the variability among these lakes in C_{pi88}

Step	F-value	Variable	r^2 -value	Model
1	4	C_{wa87}	0.78	$y = 9479 \cdot x_1 + 769$
2	4	K	0.89	$y = 9559 \cdot x_1 - 170.6 \cdot x_2 + 2524$
3	2	Open land%	0.92	$y = 9685 \cdot x_1 - 249.5 \cdot x_2 + 172 \cdot x_3 + 2804$
4	1	Total-P	0.93	$y = 9259 \cdot x_1 - 226.4 \cdot x_2 + 191.6 \cdot x_3 - 224.6 \cdot x_4 + 4939$

Other tested variables: Area of drainage area; Dm = mean depth in m; Q = mean water discharge in m^3/s ; Tw = theoretical water retention time in yr; A-areas = accumulation areas; DR = dynamic ratio; RDA = relief of drainage area; color in mg Pt/l; CaMg (hardness) in meq/l

governed by the turnover time of the lake water; i.e., the longer the water turnover time, the larger the proportion of the initial cesium load retained within the lake (mostly in the sediments).

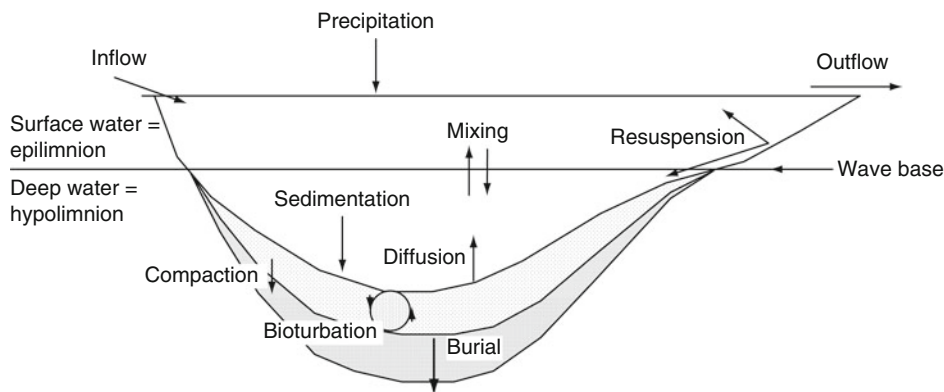
Table 2 gives statistical results, and a central part in the derivation of the following model was to account for the mechanistic principles expressed by the empirical results given in Table 2. The empirical model is static, i.e., it is only valid for the Cs-concentration in pike caught in the spring of 1988 two years after the Chernobyl fallout. The following model (see section “Dynamic Modelling”) is dynamic, i.e., it provides time-dependent predictions not just for pike, but for many species of fish with different feeding habits, for many different types of lakes, and for different fallout scenarios.

Dynamic Modelling The transport processes (sedimentation, resuspension, burial, diffusion, mixing, biouptake, etc.) illustrated in Fig. 13 are general and apply for all substances in all/most aquatic systems, but there are also substance-specific parts mainly related to the particulate fraction, the criteria for diffusion from sediments to water and the fact that some substances (e.g., nitrogen) have a gaseous phase. So, these processes have the same names for all systems and for all substances:

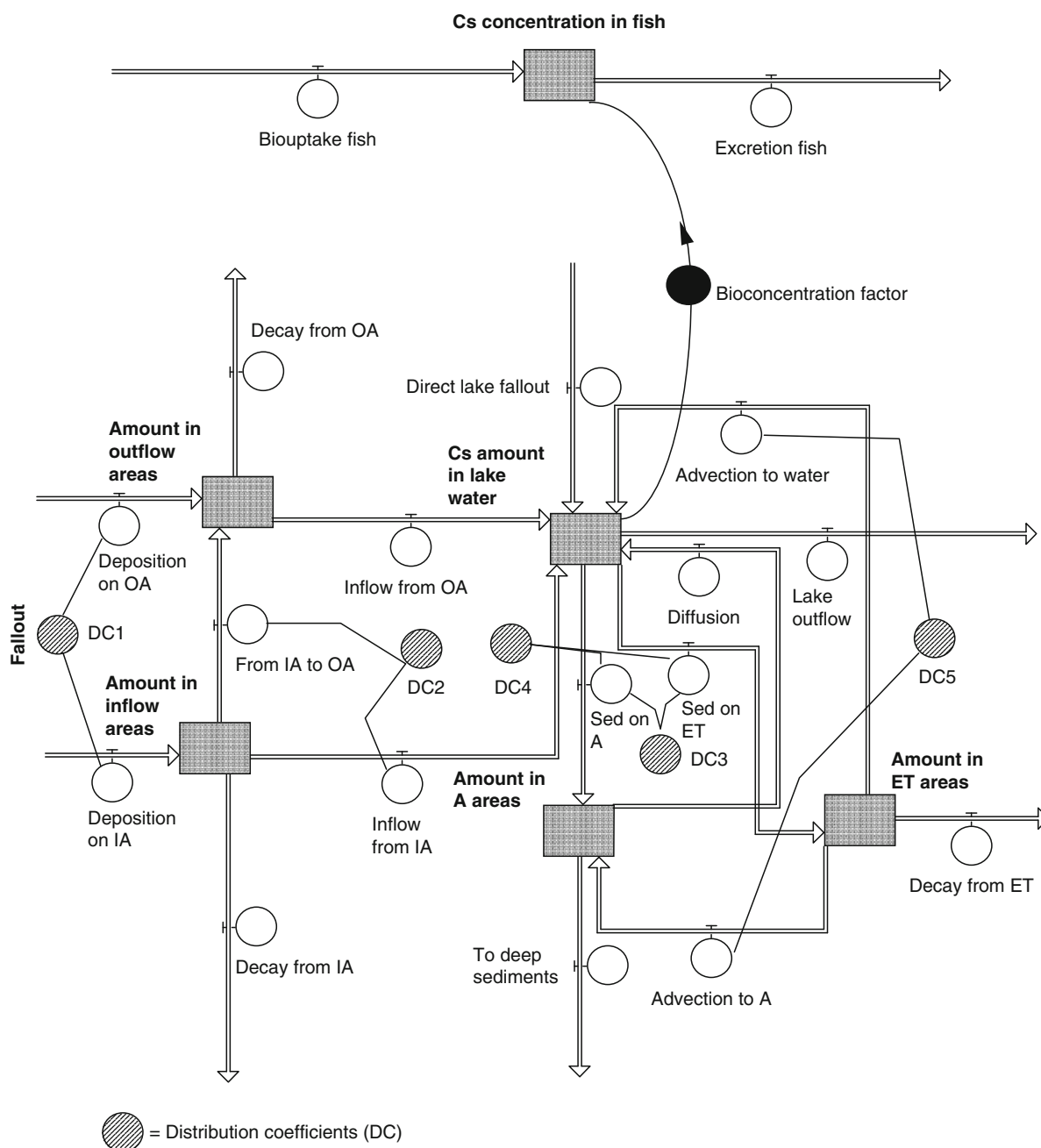
- Sedimentation is the flux from water to sediments or to deeper water layers of suspended particles and radionuclides attached to such particles.

- Resuspension is the advective flux from sediments back to water, mainly driven by wind/wave action and slope processes.
- Diffusion is the flux from sediments back to water or from water layers with high concentrations of dissolved substances to connected layers with lower concentrations. Diffusion is triggered by concentration gradients, which would often be influenced by small-scale advective processes; even after long calm periods, there are currents related to the rotation of the earth, the variations of low and high-pressures, temperature variations between day and night, etc. However, it should be noted that it is difficult to measure water velocities lower than 1–2 cm/s in natural aquatic systems.
- Mixing (or large-scale advective transport processes) is the transport between layers of water, e.g., surface-water layers and deep-water layers related to changes in stratification (variations in temperature or salinity).
- Burial is the sediment transport of matter from the biosphere to the geosphere often of matter from the technosphere.
- Outflow is the flux out of the system of water and everything dissolved and suspended in the water.

Figure 14 shows the basic structure of the model for radionuclides in focus in this section. When there is a partitioning of a flow from one compartment to two or more compartments, this is handled by a partitioning or distribution coefficient. There are five such distribution coefficients (DC) in this model:



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 13
Illustration of fundamental transport processes in lakes



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 14

Illustration of the basic structure of the mass-balance model for substances, here radiocesium in lakes

1. The DC to distribute the fallout on the catchment area to inflow and outflow areas. The catchment area is divided into two parts: “outflow areas” (\approx wetlands) dominated by a relatively fast turnover of substances and horizontal (land

overflow) transport processes and “inflow areas” (\approx dry land) dominated by vertical transport processes, first through the soil horizons, then ground water transport, and, finally, tributary transport to the lake.

2. The DC handling the partition from inflow areas to outflow areas and lake water.
3. The DC differentiating between dissolved and particulate fractions of the given radionuclide.
4. The DC regulating the sedimentation of the substance to either areas of erosion and transport (ET-areas, i.e., areas where by definition there is a discontinuous deposition of fine particles so that periods of sedimentation and resuspension occur [48] and/or to areas of accumulation (A-areas; areas with continuous sedimentation of fine particles settling according to Stokes' law; from these areas, the substance may be transported back to the water, e.g., by diffusive processes).
5. The DC describing the flux (resuspension = advection) of the substance from ET-areas back to water and/or to A-areas.

Technically, the biotic processes are treated separately from the abiotic fluxes and the two pathways are linked by means of a bioconcentration factor, which in turn is influenced by factors known to regulate the biouptake and concentration of the substance in the target organisms. In the radiocesium model, the target organism is fish eaten by man. This approach is a simplification with many advantages in predictive ecology.

In quantifying sedimentological processes, the lake is divided into four zones: surface water, deep water, ET-areas (areas where processes of fine sediment erosion and transport dominate the bottom dynamic conditions), and accumulation areas (A-areas). The following abbreviations for such fluxes are generally used rather than abbreviations in Greek letters: F for fluxes (Bq/month), C for concentrations (Bq/m³), R for rates (1/month), and M for masses (= amounts in g or Bq). Fluxes from one compartment to another are denoted, e.g., F_{SWDW} for sedimentation from surface water (SW) to deep water (DW). The resuspended particulate forms of the given radionuclide may be transported either back to the surface water (F_{ETSW}) or to the deep water (F_{ETDW}). How much will go in either direction is regulated by a distribution coefficient (DC) calculated from the form of the lake.

Sedimentation of suspended particles depends on:

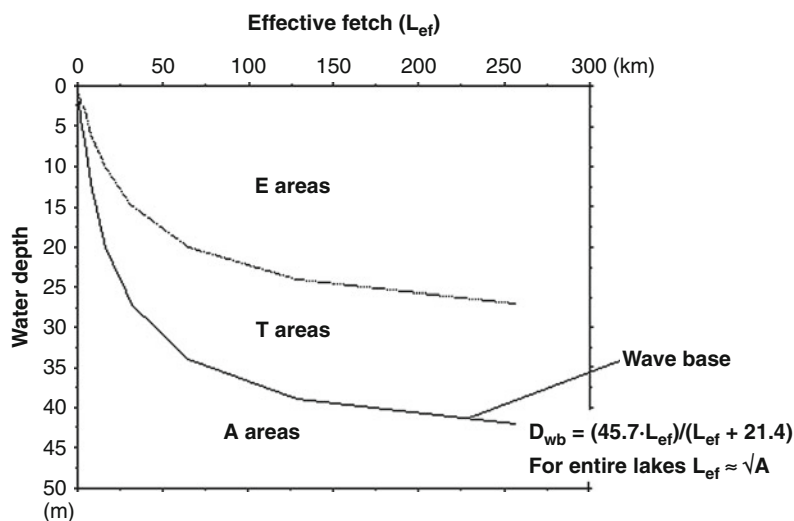
1. A default settling velocity, v_{def} which is substance-specific and related to the type of carrier particle.

Examples are 12 m/year for illite clays, the carrier particles for radiocesium [19] or 72 m/year for planktonic materials [49]. The v -value is changed into a rate (1/month) by division with the mean depth for the surface-water areas (D_{SW}) for sedimentation on the ET-areas above the wave base and by the mean depth of the deep-water areas (D_{DW}) for sedimentation in deep-water areas (the A-areas) below the wave base.

2. The concentration of suspended particulate matter (SPM in mg/l) will influence the settling velocity. The greater the aggregation of suspended particles, the bigger the flocs and the faster the settling velocity [50–52].
3. The amount of resuspended matter will settle faster than the primary materials. The resuspended particles have already been aggregated and they have also generally been influenced by benthic activities, which will create a “gluing effect.”

The wave base is often determined from the ETA diagram (Erosion–Transportation–Accumulation). This approach see (Fig. 15) focuses on the behavior of the cohesive fine materials settling according to Stokes' law (see Eq. 4). The grain size and/or the composition of the material is often used as criteria to distinguish different sediment types [53]. One can also differentiate between different sediment types by means of functional criteria (like erosion, transportation, and accumulation) of coarse sediments (friction material) or fine sediments (cohesive material). In ecological contexts, it is common to focus on the finer materials most easily set in motion/resuspension and having the highest capacity to bind pollutants [54, 55]. In defining the bottom dynamic conditions (erosion, transportation, and accumulation), this work uses the following definitions (from [56]; see Fig. 15):

- Areas of erosion (E) prevail where there is no apparent deposition of fine materials but rather a removal of such materials, e.g., in shallow areas or on slopes; E-areas are generally hard and consist of sand, gravel, consolidated clays, and/or rocks.
- Areas of transportation (T) prevail where fine materials are deposited periodically (areas of mixed sediments). This bottom type generally dominates where wind/wave action regulates the bottom dynamic conditions. It is sometimes difficult in



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 15

The ETA diagram (Erosion–Transportation–Accumulation; for more information, see [48] illustrating the relationship between effective fetch, water depth, and bottom dynamic conditions. The wave base (D_{wb}) can be used as a general criterion to differentiate between surface water and deep water in lakes

practice to separate areas of erosion from areas of transportation.

- Areas of accumulation (A) prevail where the fine materials are deposited continuously (soft bottom areas). These are the areas (the “end stations”) where high concentrations of pollutants may appear.

The generally hard or sandy sediments within the areas of erosion (E) often have a low water content, low organic content, and low concentrations of nutrients and pollutants. The conditions within the T-areas are, for natural reasons, variable, especially for the most mobile substances, such as phosphorus, manganese, iron, and radiocesium that react rapidly to alterations in the chemical “climate” (given by the redox potential) of the sediments. Fine materials may be deposited for long periods during stagnant weather conditions. In connection with a storm or a mass movement on a slope, this material may be resuspended and transported up and away, generally in the direction toward the A-areas in the deeper parts, where continuous deposition occurs. Thus, resuspension is a most natural phenomenon on T-areas.

It should also be stressed that fine materials are rarely deposited as a result of simple vertical settling

in natural aquatic environments. The horizontal velocity is generally at least ten times larger, sometimes up to 10,000 times larger, than the vertical component for fine materials or flocs that settle according to Stokes’ law [57, 58].

The particles settling on accumulation areas (A-areas) are, by definition, not further resuspended by physical forces. However, such particles with their attached radionuclides may be eaten by zoobenthos, and zoobenthos by fish, so these particles can be transported up to the lake water again in this manner. They can also be transported by migrating zooplankton between the sediments and the water column. The organic fraction of these particles could also be mineralized and decomposed by sediment-living bacteria. The newly formed dissolved phase can then migrate up to the lake water phase by diffusive processes. Bioturbation, i.e., the mechanical mixing of sediments by sediment-living organisms, can also contribute to the transport of settled particles both back to the water and downward to the biologically passive sediments.

Generally, various types of zoobenthos live in sediments down to about 5–15 cm sediment depth. This upper part of the sediment column is biologically active in the sense that the bottom fauna can influence the

physical, chemical, and biological conditions of the sediments and cause bioturbation. However, the larger animals (macro- and meiofauna) generally die if the oxygen concentration becomes lower than about 2 mg/l. This will stop the bioturbation and laminated (layered and unmixed) sediments may appear. The continuous sedimentation will cause the sediment layer to grow upward so that the bioturbation limit, i.e., the limit between the upper biological layer and the lower biopassive (or geological) layer moves upward (see [48] for further information on sedimentary processes).

Many polluting substances have a considerable particle affinity. This means that the behavior of all substances with a particular phase in aquatic systems is influenced by SPM suspended particulate matter; see (Eq. 2). In addition, the retention of a given substance in a system will be governed by the turnover time of the water; i.e., the longer the water turnover time, the larger the proportion of the initial load is retained within the system.

Geochemically, fine, cohesive materials behave differently compared to coarse (friction) materials. From the basic Stokes's equation for settling particles (Eq. 4), as well as for convenience, the limit between coarse and fine materials can be set at a particle size of medium silt.

Stokes' law expresses the settling velocity (v) as

$$v = ((d_w - d_p \cdot g \cdot d^2)/(18 \cdot AV \cdot FR)) \quad (4)$$

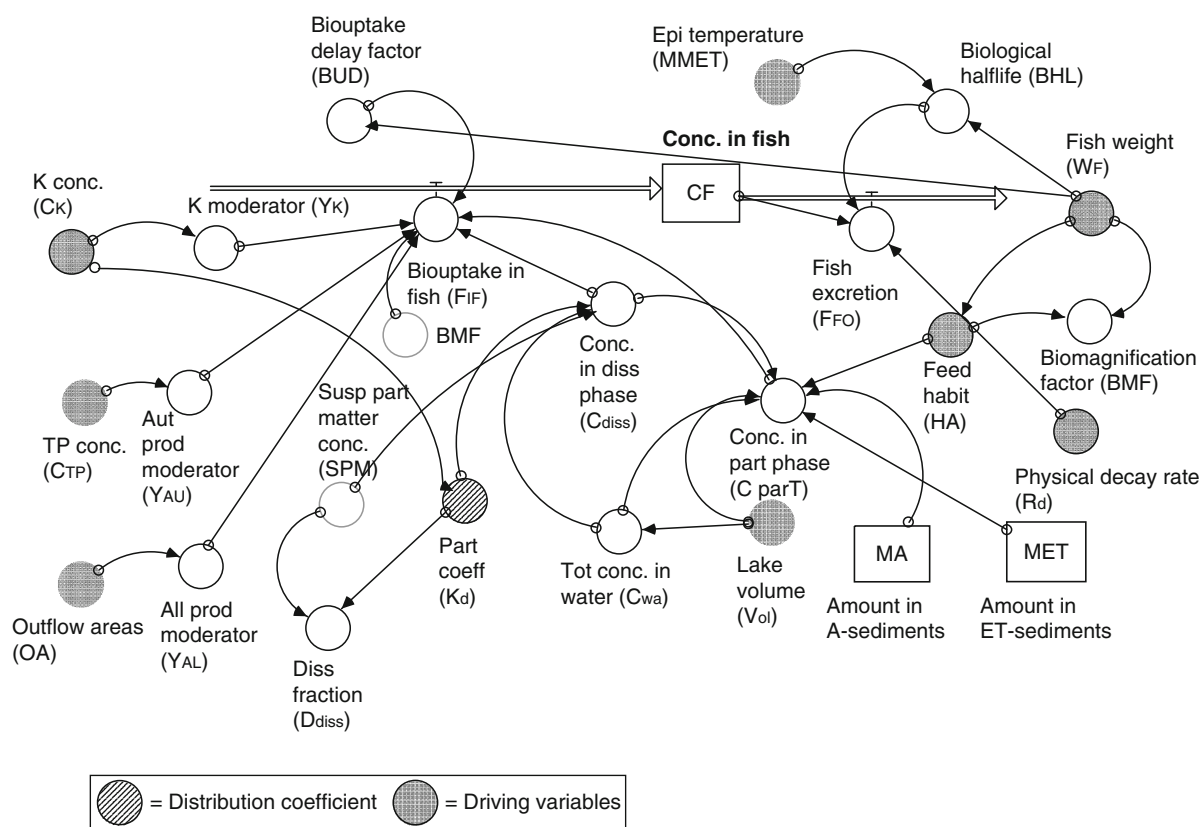
where v is the settling velocity (usually in cm/s) d_p is the particle density (usually in g d_w /cm³; g d_w = gram dry weight) d_w is the density of the lake water (often set to 1 g ww/cm³; g ww = gram wet weight) g is the acceleration due to gravity (980.6 cm/s²) d is the particle diameter (in cm) AV is the coefficient of absolute viscosity (obtained from standard tables; 0.01 P at 20°C) FR is the coefficient of form resistance (set to 1 for spheres)

The behavior of materials that follow Stokes' law (i.e., particles with a diameter between about 0.01 and 0.0001 cm) differs from the coarser friction material and the still finer material. The sedimentological behavior of the friction material is closely linked to the grain size of the individual particles [59, 60]; the sedimentological behavior of the very fine materials is

governed by Brownian motion. These latter particles are so small that they will not settle individually, but will do so if they form larger flocs or aggregates that are dense enough to settle according to Stokes' law. The cohesive materials that follow Stokes' law are very important, since they have a great affinity for pollutants. This group includes many types of detritus, humic substances, and plankton. All play significant roles in aquatic systems [35, 61].

So, the settling velocity (v) of a given particle, aggregate, or particulate pollutant, and its distribution in an aquatic system, depends on the density, size, and form of the particle (and on the hydrodynamics of the flow of water in the system). If the particle density, d_p , is close to 1, if the form factor, FR , is large, and if the diameter, d , is small, the settling velocity, v , may be very small. If v is close to 0, the particle or aggregate is conservative in the sense that it may not be deposited in the system. In many lakes (see [48]), Cl and alkalinity are often conservative substances, and color, organic matter, total-P, Si, particulate-P, and SPM are more reactive.

Modelling Biouptake in Fish To predict the concentration of radionuclides such as Cs-137 and Sr-90 in biota (primarily in fish eaten by man), a sub-model added to the main mass-balance model may be used see (Figs. 16 and 19). The sub-model discussed here has been validated (blind-tested) with good results for a great number of lakes. Some of those results are shown in Fig. 17. Empirical data are first directly compared to modelled values for Cs-concentration in 500 g pike in Lake Siggefora, Sweden (Fig. 17a). A statistical correlation of these data (22 fish) gives a r^2 -value of 0.51. This regression only covers one lake and one type of fish and, hence, has a limited range. More interestingly for a general model is to see how the model predicts for several fish species for a large number of lakes when the same model setup is used for all systems and there has been no coastal-specific "tuning" of the model variables. This is shown in Fig. 17b. The dataset consists of 185 values for seven fish species covering a very wide range of radiocesium concentrations, from 2 to over 30,000 Bq/kg ww. There are three obvious outliers. If these are omitted, the r^2 -value is 0.98 and if they are included, the r^2 -value is 0.95. The slope is very close to one. The r^2 -value for the actual



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 16

An outline of the biotic part (the sub-model for fish) of the mass-balance model for substances (here radiocesium)

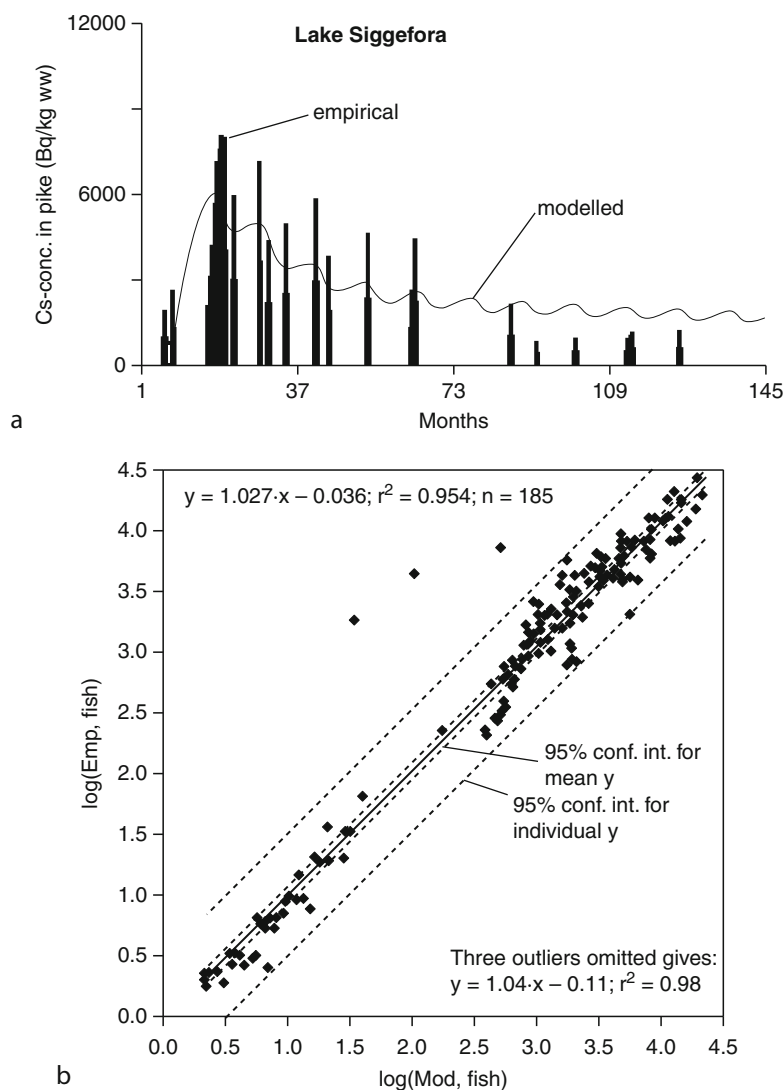
(non-log) data is 0.86. This demonstrates that the basic structure of the sub-model for Cs-137 in biota (fish) in lakes is good. If such good predictions can be obtained for Cs-137 for fish in lakes, this holds great promise also for coastal areas and other types of contaminants in future models.

Basically, the fish model relates the concentrations of a given radionuclide in water, on suspended particles, and in sediments to concentrations in fish. In the model, the bioconcentration factor (BCF; see, e.g., [62]) is influenced by factors known to change the biouptake and retention of radionuclides in biota. Algorithms for this are included in the sub-model for fish and are only briefly described below:

- The concentration of the given radionuclide in dissolved and particulate phases: The concentration in dissolved form will mainly influence planktivorous fish, since they feed on plankton,

which are susceptible to the dissolved part of the radionuclide in the lake water; the radionuclide concentration in sediments mainly affects benthivores.

- The feeding habits of the fish (a general algorithm is used for planktivores (eating plankton), piscivores (eating fish), omnivores (eating “everything”), and benthivores (eating bottom fauna), but not for herbivores (eating plants) due to lack of data for model tests for such fish.
- The weight of the fish (a general algorithm where the weight of the fish influences the response time related to biouptake [BUD in Fig. 16] and the biological retention time of the radionuclide in the fish).
- The trophic characteristics of the coastal system: A general algorithm accounting for biological dilution, i.e., the fact that the mean concentration in fish basically is basically defined by the total amount



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 17

(a) The relationship between empirical data and modelled values for radiocesium in 500 g pike in Lake Siggefora, Sweden and (b) results of model validation for lake fish from 185 samples from different types of fish from European lakes (Modified from [19])

of the radionuclide in the fish divided by the total biomass of the fish; if the amount of the radionuclide is constant, the concentration will decrease if the fish biomass increases, and vice versa.

- **Water temperature:** A general algorithm is used where the water temperature influences the biological half-life of the radionuclide in fish, and hence also the excretion of the radionuclide from fish. This model uses an algorithm for the biological
- **Chemical dilution:** If the water contains a large amount of substances that are taken up by biota in a way similar to the given radionuclide, the biouptake of the radionuclide will be lower. The K-concentration in water specifically influences

half-life that has been tested for water-soluble substances (such as radionuclides), but not for lipophilic substances (such as DDT and PCB), so this approach is substance-specific.

the biouptake of Cs-137 and the Ca-concentration specifically the biouptake of Sr-90. There are substance-specific algorithms for this in the model.

So, technically, the biotic processes are treated separately from the biotic fluxes and the two pathways are linked by means of a bioconcentration factor see (Fig. 14). This is a simplification with many advantages in predictive modelling:

1. One can then, as already stressed, predict the target concentrations in biota directly without data on biomasses, which are generally extremely hard to access and predict. This model uses compartments of one mass unit: Cs-137 is transported from one mass unit of lake water to one mass unit of fish. This means that the model automatically yields concentrations of the given substances in fish. Most mass-balance models use the entire lake or a given part of the lake as the unit of size and calculate the concentration of a substance in water as the ratio between the total amount (g or Bq) of material water related to in the total volume of water (l or m³).
2. One can relatively easy test (calibrate and/or validate) models of this type, since concentrations in biota are easily measured.

This model structure generally gives better predictive success than traditional models using connected or parallel biotic compartments (like phytoplankton, zooplankton, zoobenthos, prey fish, and predator fish models of the type illustrated in Fig. 11a). The uptake of the given substance from water to biota is still modelled dynamically.

These factors are accounted for in this model see (Fig. 16). Such approaches are most relevant in cases when the biotic fluxes of the given substance are much smaller than the abiotic fluxes. This is the case for practically all metals and radionuclides in most lakes (see [63]).

The differential equation for the radionuclide concentration in fish (i.e., C_F in Bq Cs-137 per kg wet weight of fish) is given by

$$C_F(t) = C_F(t - dt) + (F_{IF} - F_{FO}) \cdot d \quad (5)$$

where

F_{IF} is the biouptake (Bq/month·kg)

F_{FO} is the excretion (Bq/month·kg)

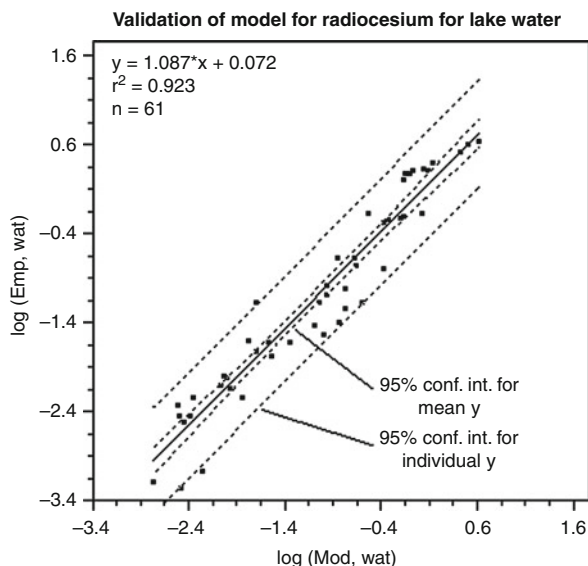
Equation 5 is simple but the model is relatively comprehensive to achieve generality. The number of obligatory driving variables is still small and they are all readily accessible from standard maps and/or monitoring programs (an important demand for the practical utility of the model).

Model Testing The lake model for radiocesium has been validated for a broad set of independent lakes for two target variables, [1] radiocesium concentrations in lake water and [2] radiocesium concentrations in fish see (Fig. 17b). The validation lakes cover a wide range of lake characteristics, although they are all European: sizes are from 0.041 to 1,147 km², mean depths are from 2.9 to 90 m, characteristic pHs are from 5.1 to 8.5, trophic levels range from oligotrophic (phosphorus concentrations from 10 µg/l) to highly eutrophic (phosphorus concentrations = 60 µg/l), fallout ranges from 2 to 130 kBq/m².

The validations were done in the following manner. Empirical data were first directly compared to modelled values for all lakes individually according to the procedure illustrated in Fig. 17a. All data-series like the one given in Fig. 17a have been compiled for all the validation lakes. The results for the Cs-concentrations in lake water are given in Fig. 18. The r^2 -value is 0.92. The figure also gives the 95% confidence intervals for the mean y and the individual y and the regression line. The slope is also almost perfect, 1.087. This regression is based on 61 individual data covering a range from 0.001 to 4.5 Bq/l. This is, of course, a very good result for a validation.

Figure 17b gave the corresponding results for Cs-concentrations in lake fish.

Figure 19 gives the fallout map showing calculated deposition of radiocesium (kBq/m²) over Sweden from the Chernobyl accident. The figure also shows the boundaries of two provinces (Gäveborgs and Västernorrlands counties) in central Sweden where intensive studies and full-scale experiments were carried after the Chernobyl accident to speed up the recovery related to radiocesium concentrations in lake fish. The map to the right in Fig. 19 shows radiocesium concentrations in lake fish (values in kBq/kg wet substance in one 100 g perch caught in the fall of 1987; small perch is the most abundant fish in Swedish lakes) based on empirical data from 644 lakes. The interesting



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 18

Validation of the radiocesium model for lake water
(From [19])

aspect concerns the differences between the fallout map and the map showing concentrations in fish and how to explain this difference. The model discussed here can statistically and causally explain about 90% of this difference, as shown in Fig. 17b.

With this dynamic modelling approach, it is also possible to make realistic predictions on how the situation would likely change in the future – e.g., for how long there will be a radiocesium problem with concentrations of Cs-137 higher than 1,500 Bq/kg in fish in Swedish lakes (the guideline value for consumption and sale) from the Chernobyl accident. Figure 20 gives two maps on modelled radiocesium levels in pike (a top predator). In 1987, 1 year after the Chernobyl accident, there were about 14,000 lakes in Sweden with fish having more than 1,500 Bq/kg. From Fig. 20, one can note that on a regional basis all lakes in Sweden should have fish with radiocesium levels below 1,500 Bq/kg in 2010.

It is often argued that the quality of Science is related to the possibilities to make meaningful predictions. This argument would give Mathematics and Physics a pool position in the scientific community. It has long been argued that due to the complex nature of aquatic ecosystems, it will never be possible to predict

important target variables, especially not with more comprehensive dynamic models. Those arguments are no longer valid. The key lies in the structuring of the predictive models.

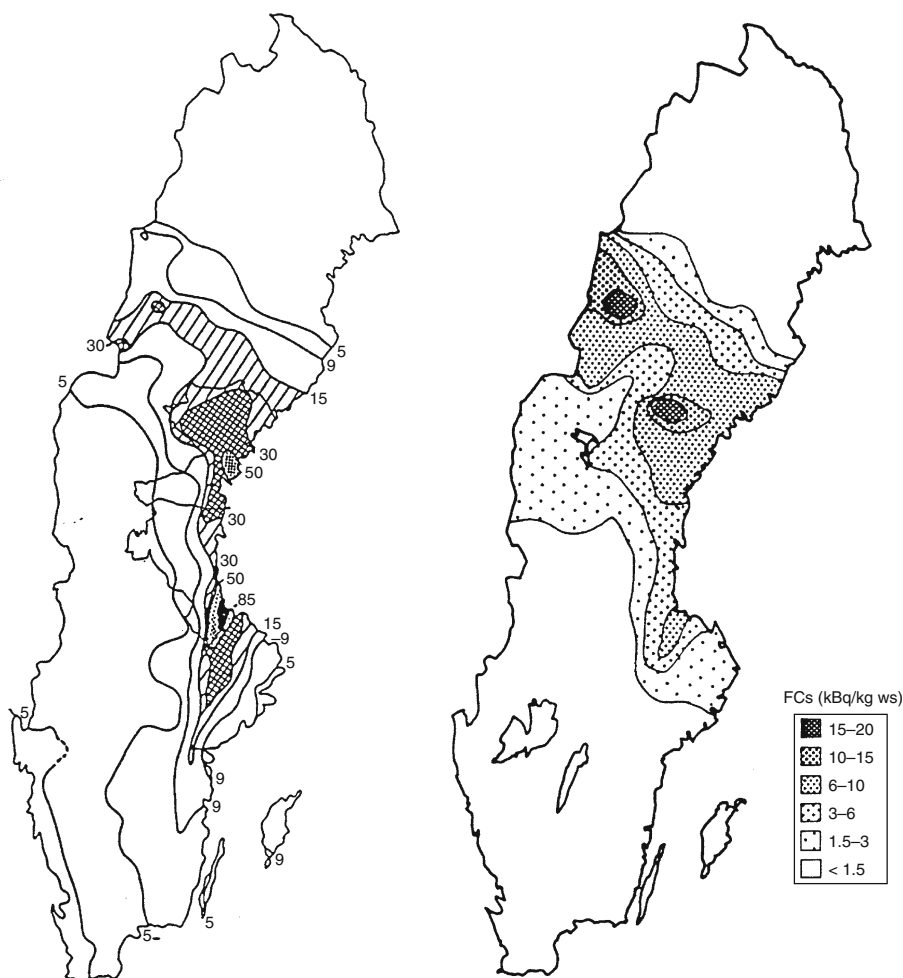
Modelling Coastal Marine Systems

Introduction, Background, and Aim Coastal areas are target areas for research and environmental management for several reasons:

- The coastal zone may be regarded as a “pantry and a nursery” for the sea and coastal areas can have a very high bioproduction (20 times higher than the most productive areas on land; see [64]). All three functional groups of primary producers, phytoplankton, benthic algae, and macrophytes are present in coastal areas, but not in open water areas. Where there is a high primary production, there is also a high secondary production of zooplankton, zoobenthos, and fish (see [49, 65]).
- The coastal zone is also a zone of conflicts where many different users, such as professional and leisure-time fishermen and people responsible for recreation, shipping, and environmental management and research, place different demands on the coastal waters and apply different criteria to set the value of coastal waters and to define desired conditions (see [49, 65]).
- The coastal zone is also the “recipient” of many types of pollutants, such as radionuclides, nutrients, metals, and organic toxins (see [69, 70]).

Due to the hydrodynamic conditions characterizing a given coastal area, such as the Coriolis-driven coastal jet-zone, tidal effects, and winds, the transport of water pollutants along coasts, the biological value of the coastal zone, and the different demands on the utilization of coastal waters, it is easy to understand why so much interest and research concern this zone. However, in contexts of aquatic radioecology, it is easy to conclude that much more effort has been focused on lakes (see, e.g., [1, 71–74]) and rivers (see, e.g., [75–78]) than on coastal areas. In fact, it is very difficult to find the radionuclide data necessary to test coastal models.

The basic aim of this work is to present a generic coastal model. It is meant to be valid for most coastal



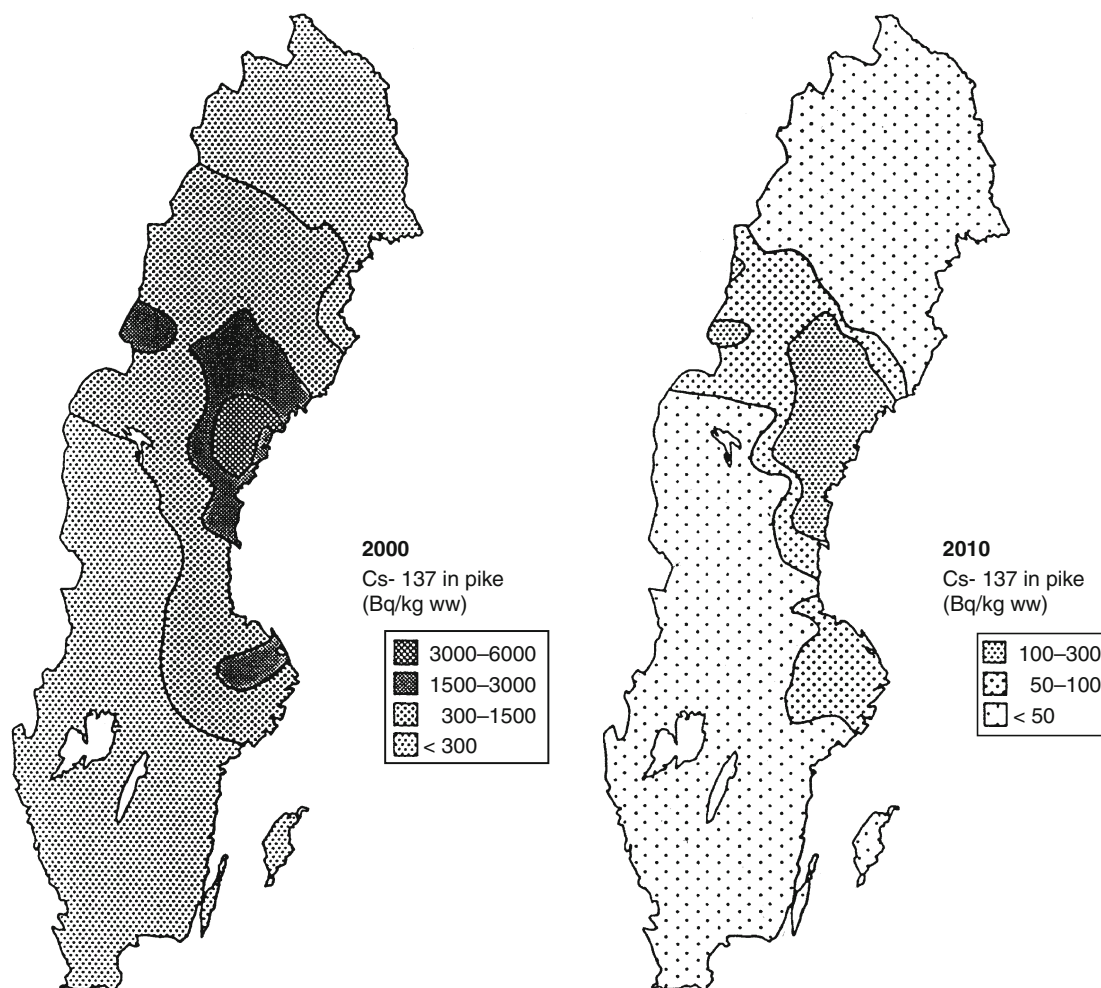
Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 19

Left; fallout map showing the deposition of radiocesium (kBq/m^2) over Sweden from the Chernobyl accident. *Right*; map showing radiocesium concentrations in lake fish (values in kBq/kg wet substance in 1 hg perch caught in the fall of 1987) based on empirical data from 644 Swedish lakes

areas and for most substances, not only radionuclides. It is dynamic and based on ordinary differential equations (and not on partial differential equations). It is intended to give seasonal (monthly) variations on radionuclide concentrations in water and sediments and not variations at finer temporal or spatial scales. It is meant to be practically useful in radioecology, so the obligatory driving variables should be easily accessed from standard bathymetric maps and monitoring programs. Basically, this model is intended to be analogous in terms of structure and applicability to

lake and river models used in the MOIRA decision support system (DSS; see [79]).

In radioecology, there are some coastal models, e.g., Threetoxt and Coasttoxt. Unlike the model discussed in this work, these models are distributed 2-d models based on partial differential equations; the model POSEIDON is based on several interlinked boxes. These models are used in the RODOS DSS (see <http://www.rodos.fzk.de/>) and they are mainly designed to handle short-term (hours to days) spatial variations after accidental fallouts. They are driven by online



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 20

Modelled radiocesium levels in pike in Swedish lakes in the years 2000 and 2010

meteorological data (winds, temperature, and precipitation) and operate over much smaller temporal and spatial scales than the coastal model discussed in this context.

The ideal dataset for coastal modelling would have reliable measured data on several radionuclides and their concentrations in coastal water, sediments, rivers, and the sea outside the coast, as well as relevant co-parameters describing morphometry, water quality, etc., for several coastal areas covering a wide geographical domain. While waiting for that ideal dataset, this section may be regarded as “stepping-stone” for future model validations in coastal radioecology.

The following specific questions will be discussed in this section:

- Which fluxes (Bq/month) to, within, and from a given coastal area can be expected to be large and small? Evidently, it is more important to predict large fluxes with a smaller uncertainty than smaller fluxes. The answer to this question is also fundamental for sensitivity and uncertainty analyses (see [63]).
- The exchange processes between the coast and the sea are often of fundamental importance for the concentrations of radionuclides (and other substances as well) in coastal water, sediments, and biota (see, e.g., 19). For which coast types will the concentration in the coastal water be close to the concentrations in the sea and when and why will there be differences?

- In models of this kind (i.e., process-oriented mass-balance models based on ordinary differential equations and monthly timescales) there are always uncertainties related to the model structure and uncertainties in driving variables – but which uncertainties related to fluxes and driving variables are generally small, and which are important and need to be reduced to increase the predictive power of the model?
- In this model, there are generic parts that apply for all substances/radionuclides and substance-specific components. In applying models of this kind for different radionuclides, it is essential to stress this point. This may be done by using the algorithm for sedimentation as an example. Sedimentation is a general process that applies for all substances with particulate phases. Sedimentation is given in this coastal model by the following algorithm, which gives sedimentation from surface water (SW) to deep water (DW) separated by the wave base:

$$F_{\text{SWDW}} = M_{\text{SW}} \cdot (v/D_{\text{SW}}) \cdot \text{PF} \cdot (1 - \text{ET}) \cdot [(1 - \text{DC}_{\text{res}}) + 10 \cdot \text{DC}_{\text{res}}] \quad (6)$$

M_{SW} is the mass (amount) of the radionuclide/substance in the SW compartment in Bq.

v is the settling velocity of the particulate fraction (PF) of the radionuclide in m/month. The v -value used here is valid for suspended particulate matter (SPM).

D_{SW} is the mean depth of the surface-water compartment in the given coastal area in m.

ET is the fraction of ET-areas (i.e., areas where discontinuous sedimentation of fine particles appear) in the coastal area; these are the areas above the wave base (WB); the deep-water areas are the areas below this water depth (ET is dimensionless).

DC_{res} is the fraction of resuspended particles of the total amount of particles in the surface water compartment; resuspended particles from the ET-areas have a higher settling velocity (a factor of 10 compared to the primary particles) and a shorter distance to settle. C_{res} is calculated automatically in the model.

The only substance-specific part in Eq. 6 is the particulate fraction (PF). The dissolved fraction DF is $(1 - \text{PF})$; DF is about 0.95 for Sr-90, about 0.5 for Cs-137 and about 0.05 for Pb-210 in aquatic environments

(see [1, 80] and Fig. 8). Radiostrontium is, as already stressed, more mobile than radiocesium and this is expressed by the substance-specific DF-value. From this, the following question will be addressed in this section: How important is the DF-value for predictions of radionuclide concentrations in coastal waters, sediments, and biota?

Other interesting questions that may be addressed using this model are:

- How will coastal morphometry (coastal size and depth conditions and the openness toward the sea) influence concentrations in water, sediments, and biota? Is it possible to find some general guidelines related to the vulnerability of coastal areas to radionuclide contamination?
- How would tidal effects influence radionuclide concentrations in coastal areas?
- If a coastal area is contaminated by radionuclides, can the model be used to simulate realistic responses and feasible remedial measures?

It should be noted that it would be extremely difficult and costly to scrutinize questions like these using more traditional approaches based on field studies in coastal areas. This is an area where process-based modelling can provide an important input to clarify what is happening in natural coastal areas.

To address such questions, this entry is structured in the following way: Section [The Coastal Model](#) gives a short model description, Section [A Ranking of Fluxes](#) gives a ranking of key fluxes for a defined coastal area in the Baltic Sea, Section [Key Factors Influencing the Ranking of Fluxes and Concentrations](#) discusses factors influencing the ranking of fluxes and concentrations and Section [Remedial Measures Simulations: A dredging Scenario](#) gives a remediation scenario on dredging.

First, however, Section [Defining Coastal Area Boundaries](#) will discuss criteria to define the fundamental unit, the coastal area. It is evidently easy to define what is a lake. A river stretch may be defined by the area between two inflowing tributaries. But where and how should one draw the borderlines toward the sea and/or adjacent coastal areas? If such borderlines are drawn arbitrarily, one would get arbitrary volumes, areas, and mean depths and the mass-balance model would lose predictive power (see [19]).

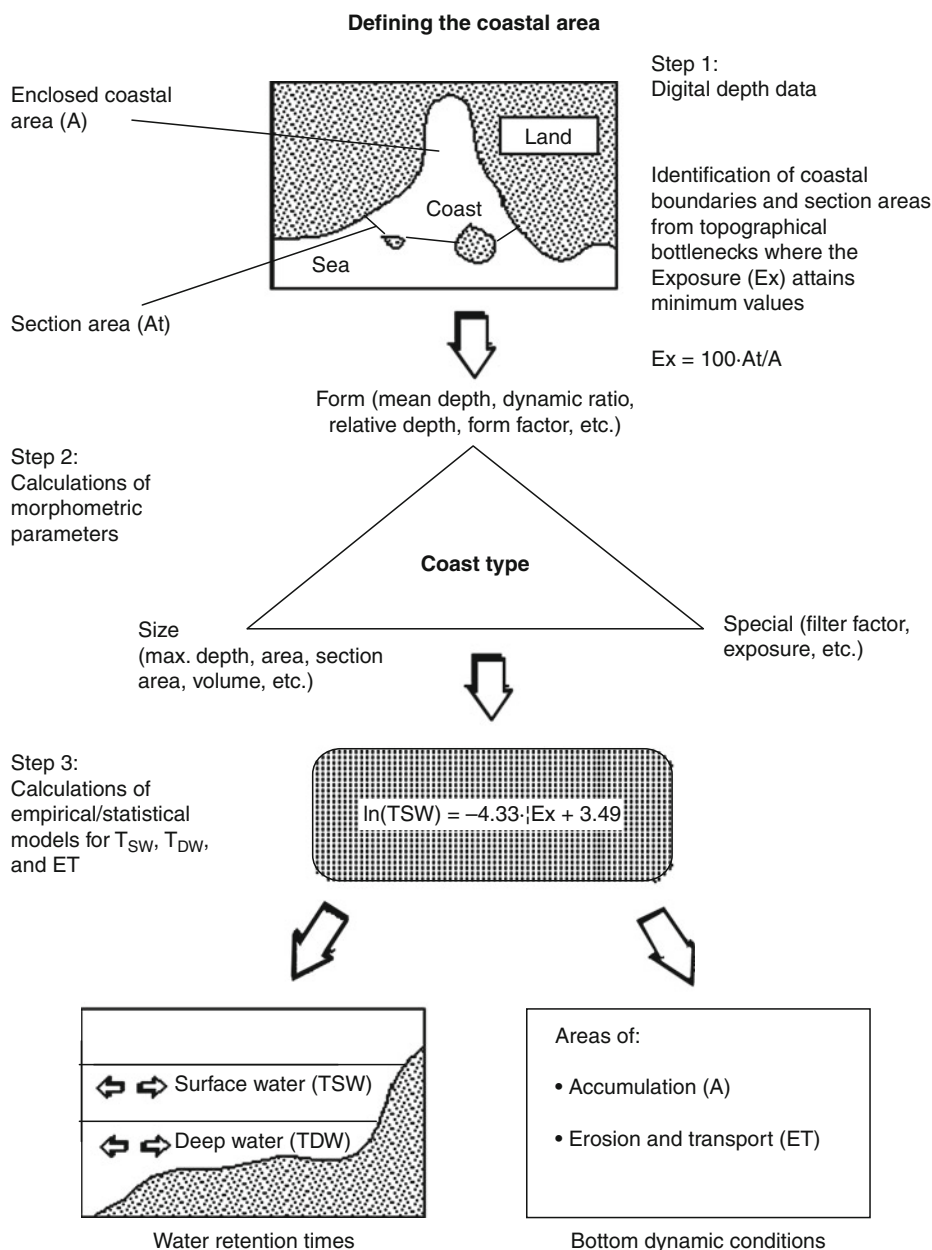
Secondly, fundamental criteria for the water exchange between the coast and the sea will be briefly discussed based on three coast types with different characteristics. The main model in this section is the coastal mass-balance model handling internal fluxes. This model is a modified version of a model for suspended particulate matter (SPM) presented by [81]. It will be outlined in the following text.

Defining Coastal Area Boundaries Today, there are many modern digitized methods applicable in aquatic studies including Geographical Information Systems (GIS). Those methods enable quick and effective estimations of morphometric variables (see [82]). Special considerations should be given to define the boundaries of coastal areas. After digitizing and interpolation, the area outside the coastal area can be masked, resulting in one vector file containing coordinates for all shore lengths (of islands and the coastline) and openings to the surrounding open water areas, and one raster file containing the pixel size (header) and the digital elevation model of the coastal area. Figure 21 illustrates steps to define a coastal area and the boundaries to the open water areas or adjacent coastal areas at the topographical bottlenecks using GIS-techniques where the exposure ($Ex = 100 \cdot At/A$) attains minimum values when different alternatives for settling the boundary lines are tested (At = section area; A = enclosed coastal area; method from 81). Once the coastal area is defined, one can also define important variables for mass-balance calculations, such as the coastal volume (regulating the concentration of any given substance), and important morphometric parameters for internal fluxes, such as the mean depth and the water surface area, and key variables regulating the water exchange between the coast and the sea, such as the section area, the exposure, and the filter factor see (Fig. 22). This method of defining coastal areas also opens a possibility to use empirical models to estimate, e.g., the theoretical water retention times of the surface water see (Fig. 23) and the deep water, and the bottom dynamic conditions (regulating sedimentation, resuspension, and diffusion) from morphometrical parameters (such as area, mean depth, and section area).

Water Exchange in Different Coast Types Figure 23 shows an overview of the three main coast types, (1) tidal coasts, (2) open coasts, and (3) archipelago coasts and bays, and how the theoretical surface water retention time (T_{SW}) may be calculated for each type and for any given coast that may be a combination of all three (for further details also on these matters, see Table 3 and [19]). From the theoretical surface water retention time (T_{SW} in months), one can calculate the surface water flux (Q_{SW} in $m^3/month$) as V_{SW}/T_{SW} , where V_{SW} is the surface water volume (m^3) and the flux of the given substance, $Q_{SW} \cdot C_{sea}$ in $g/month$, where C_{sea} is the concentration of the substance in the sea outside the given coastal area (g/m^3).

As a rule of thumb, one can say that the costs of establishing an empirical value of the theoretical water retention time from traditional field measurements (using dye, current meters, *etc.*) is about US \$30,000 for one coastal area in the size range from 1 to 100 km^2 (see [19]). In many contexts, it may not be very meaningful to build a model if it is a prerequisite that such field work first must be carried out to determine the water retention time as a driving variable (x) predicting target y -variables, such as concentrations of radionuclides in water and fish. This means that it is of major importance that water retention time can, in fact, for many coastal areas be predicted very easily from one coastal morphometric variable, i.e., the exposure (Ex).

The Coastal Model The coastal model see (Table 4 for default values for the obligatory driving variables and some calculated model constants) is modified from two validated coastal models (for suspended particulate matter and phosphorus) based on the same modelling principles and structures as discussed for the lake model in this entry. One can note that for all the tested coastal areas, the model, as applied for predicting phosphorus concentrations (see [83]), generally predicts within the uncertainty bands given by the variations in the empirical data. Sensitivity and uncertainty tests of the model have also been presented (see [81]). The model can be used for coastal areas of the types shown in Fig. 23 and combinations of these three coastal types: coasts where the water exchange is dominated by tides, open coasts dominated where the water renewal is governed by coastal current activities, and archipelago coasts. In estuaries (with freshwater inflow), one must



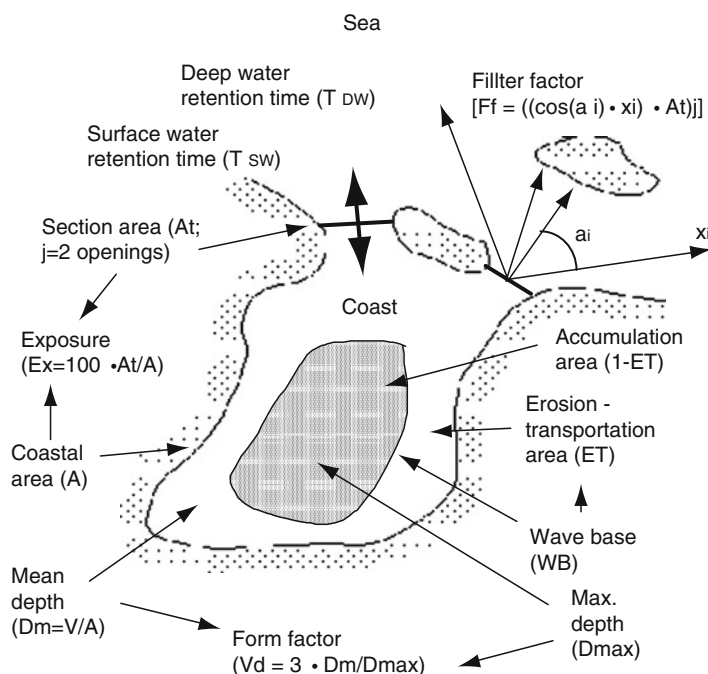
Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 21

Illustration of the steps to define the boundary lines for a coastal area (step 1) and how to calculate important coastal parameters in contexts of mass-balance modelling at the ecosystem scale, i.e., for defined entire coastal areas (steps 2 and 3).

also evidently account for the inflow of water (and radionuclides) from all major tributaries to the coastal area.

An important demand, related to the practical utility of the model, is that it should be driven by variables

readily accessed from standard monitoring programs or maps. The obligatory driving variables include four morphometric parameters (coastal area, section area, mean and maximum depth; see Fig. 22), latitude (to predict surface-water and deep-water temperatures,



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 22

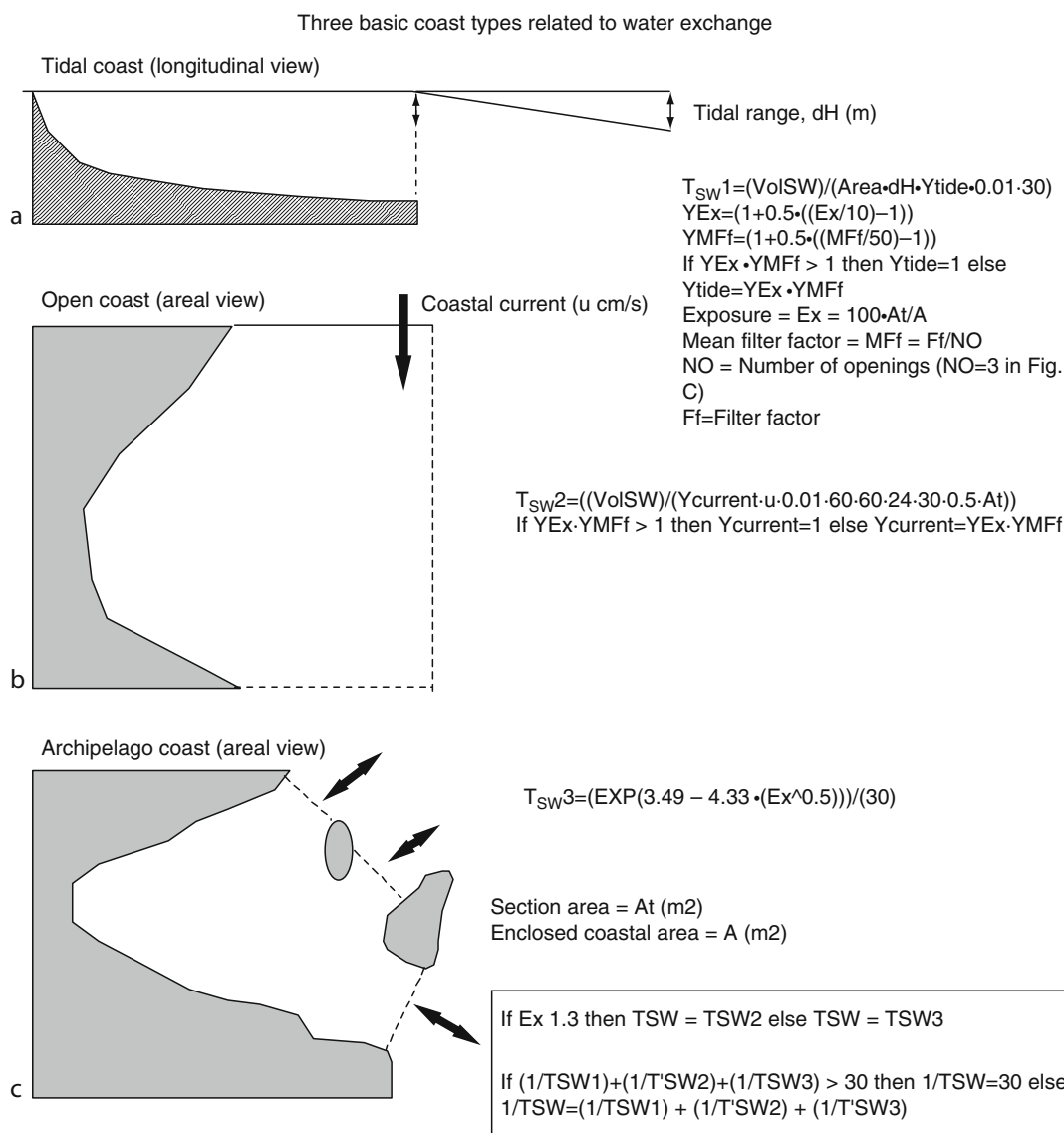
Illustration of key coastal parameters in mass-balance modelling that can be determined quickly from digitized bathymetric maps using GIS-methods (*GIS* Geographical Information System). The most important criteria in this context is to define the boundary lines, i.e., where the coastal area ends and the sea or adjacent coastal area begins. The approach illustrated in this figure is to define the boundary lines so that the topographical openness (the exposure, Ex , defined by the ratio between the section area, A_t , and the enclosed coastal area, A) attains a minimum value. This is the “topographical bottleneck” method. The filter factor (a measure similar to the fetch) describes how islands and topographical barriers between the defined coastal area and the sea act as energy filter for the wave impact on the coastal area. In this figure, the filter factor is illustrated for one of the two section areas in this coastal area (a_i = the angle between two radials sent out from the opening over open water; x_i = the length of each radial; i = the number of radials at each opening; j = the number of openings; $j = 2$ in this example)

stratification, and mixing) and the concentration of the given radionuclide in the sea outside the given coastal area, which is estimated in this entry using a simple approach based on the ecological half-life.

Like the lake model, the coastal model has four compartments. Two water compartments, surface water and deep water. The separation between these two compartments is done not in the traditional manner from measured water temperatures but from sedimentological criteria, as the water depth separating transportation areas from accumulation areas (the wave base). The model also has two sediment compartments, the ET-areas, i.e., the erosion and transportation areas where fine sediments are discontinuously

being deposited, and the A-areas, i.e., the accumulation areas where fine sediments are continuously being deposited. The processes accounted for are inflow and outflow from the outside sea (and for estuaries also from rivers) via surface and deep water, direct fallout onto the water surface of the coastal area, sedimentation, burial (the transport from surficial A-sediments to underlying sediments), resuspension, diffusion, and mixing between surface and deep water.

A Ranking of Fluxes A ranking of all fluxes to, within and from a “typical” Baltic coastal area, defined here as the default area see (Table 4), is exemplified in Fig. 24 (see [20]). One can note that the fluxes in and out from



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 23

Illustration of the three main coast types and how the theoretical surface water exchange (T_{SW}) may be estimated for each type

the sea are the most dominating fluxes and that burial, diffusion, sedimentation to the deep water, and resuspension to the deep water are the four smallest fluxes. These results are valid for the default coastal area (but not necessarily for other coastal areas) and they imply that for this area, the conditions in the sea are of utmost importance for the conditions prevailing in the given coastal area. Figure 25 compares calculated concentrations in river water (Fig. 25a, curve 1), in coastal

water (Fig. 25a, curve 2), and in the sea (Fig. 25a, curve 3), in accumulation area sediments (A-sediments; Fig. 25b) and concentrations in particulate and dissolved phases. From Fig. 25, one can conclude that the short-term changes in concentrations in coastal water the month after the fallout are very significant compared to the long-term changes after the initial month(s) when the water concentrations in the coastal area approach the concentrations in the sea.

Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Table 3 Empirical models used in this dynamic coastal model to predict theoretical surface water and deep-water retention times (T_{SW} and T_{DW} in days) and the fraction of ET-areas (ET) from morphometric parameters in archipelago areas and bays. Note that if reliable empirical data are at hand for a given coastal area on T_{SW} , T_{DW} , or ET, such data should be used rather than the values predicted by these empirical models (From [81])

Regression	r^2	Number of coastal areas
$\ln(T_{SW}) = (-4.33(\sqrt{Ex}) + 3.49)$	0.95	14
$T_{DW} = (-251 - 138 \log(At) + 269 \log(Vd))$	0.79	15
$A = 1 - ET = (D_{max} - D_{TA}) / (D_{max} + D_{TA} \cdot \text{EXP}(3 - Vd \wedge 1.5)) \wedge (0.5 Vd)$ $D_{TA} = WB = Y_{Ex1} (45.7(\text{Area } 10 \wedge (-6)) \wedge 0.5 / (21.4 + (\text{Area } 10 \wedge (-6)) \wedge 0.5))$ If $D_{TA} \geq D_{max}$ then D_{max} else D_{TA} If $Ex < 0.003$ then $Y_{Ex1} = 1$ else $Y_{Ex1} = (Ex/0.003) \wedge 0.25$		

Model domain: $0.002 < Ex < 1.3$; $0.0006 < At < 0.08$; $0.5 < Vd < 1.5$; data from coastal areas with very little tidal range; note that T_{SW} and T_{DW} are never permitted to be < 1 day and T_{DW} never > 120 days
 Ex (exposure) = $100 \cdot At / A$; Vd (the form factor) = $3 \cdot D_m / D_{max}$; At = section area (km^2); A = coastal area (km^2); D_m = mean coastal depth (m); D_{max} = max. coastal depth (m); A = the fraction of A-areas; ET = the fraction of ET-areas; $D_{TA} = WB$ = the wave base (m)

Key Factors Influencing the Ranking of Fluxes and Concentrations Figure 26 shows how the conditions in the sea influence the calculated concentrations of radiocesium in coastal water under different presuppositions. In Fig. 26a, the ecological half-life regulating the concentration in the sea has been set to 1, 3, and 5 years. This will strongly influence the long-term, but not the short-term, concentrations in the coastal area. Fig. 26b illustrates how three different values of the theoretical surface water retention time (T_{SW}), which regulates the transport between the sea and the coastal area ($Q_{SW} = V_{SW}/T_{SW}$), will influence the calculated water concentrations of radiocesium in the default coastal area. This T_{SW} -value will regulate the “mid-term” conditions more than the short-term and the long-term conditions.

From these results, it is evident that for the default coastal area in the Baltic Sea with a very small tidal amplitude some fluxes and model uncertainties are relatively unimportant. Figure 27 shows that the value used for the particulate fraction (PT set to 0.05, 0.5 and 0.95) is very important for the concentrations in dissolved and particulate phases (Fig. 27a), for the

concentration in A-sediments (Fig. 27b) and for the concentrations in fish (Fig. 27c; here calculated for 20 g planktivores; results from [20]). If the PF-value is low and the dissolved fraction high, sedimentation is low and the concentration in the dissolved phase and in small planktivores fish high; and vice versa if the PF-value is high.

Since resuspension, deep-water fluxes, diffusion, and burial are not very important in the default coastal area, the model predictions for the concentrations of radiocesium in water are not very sensitive to the value used for the ET-areas, the theoretical deep-water retention time (T_{DW}), the diffusion rate (R_{diff}), or the age of the A-sediments regulating burial (T_A).

An important point here is that using a process-based model, one can also identify coastal areas where the conclusions for the default coastal area in the Baltic Sea are not valid.

Remedial Measures Simulations: A Dredging Scenario Simulations for one of the few feasible remedial measures for radionuclide contamination of coastal areas, fertilization, have been presented by [19].

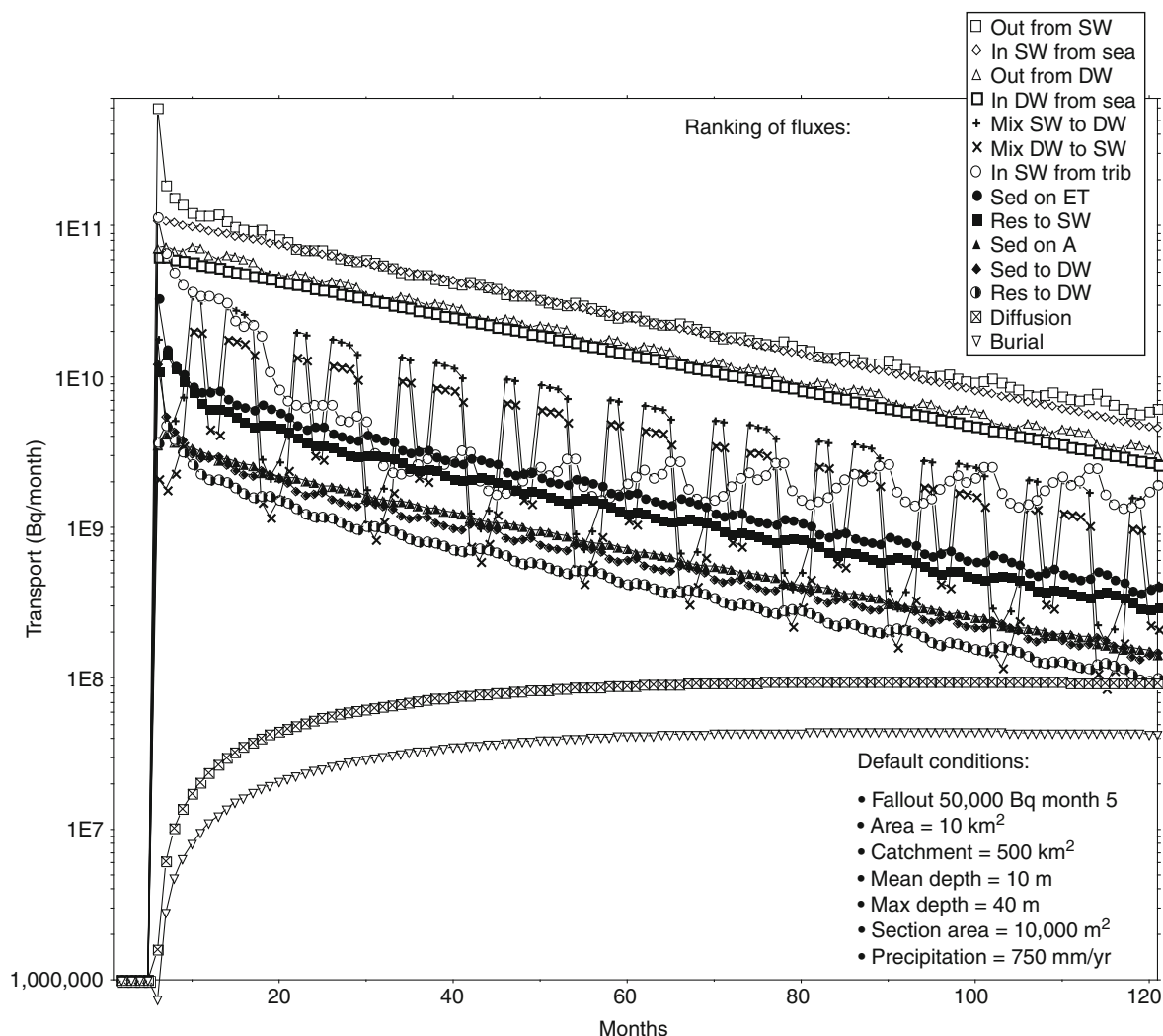
Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Table 4 Default values for the obligatory coast-specific driving variables, some model variables and calculated values for some model constants. The default values have been selected to represent a “typical” Baltic coastal area

<i>Obligatory coastal-area-specific driving variables</i>
Coastal area (= Area) = $10 \cdot 10^6 \text{ m}^2$
Latitude (= Lat) = 60°N
Max depth (= D_{\max}) = 40 m
Mean depth (= D_m) = 10 m
Mean annual precipitation (= Prec) = 750 mm/year
Salinity (= Sal) = 7‰
Section area (= At) = $10,000 \text{ m}^2$
<i>Generic default values for some model variables</i>
Age of ET-sediments = 1 month
Organic content of A-sediments (= IG) = 10% dw
Thickness of A-sediment layer = 10 cm
Water content of A-sediments (= W) = 75% ww
Bulk density of A-sediments = 1.26 g/cm^3
Potassium concentration = 1.75 mg/l
<i>Calculated values for the default coastal area for some model constants</i>
Duration of growing season (T_{dur}) = 189 days
ET = 0.73 (the ET-areas constitute 73% of the coastal area)
Exposure (Ex) = 0.1
Form factor (Vd) = 0.75
Mean depth of the SW compartment, $D_{\text{SW}} = 7.07 \text{ m}$ (from the water surface to the wave base)
Mean depth of the DW compartment, $D_{\text{DW}} = 12.93 \text{ m}$ (from the wave base to the maximum depth)
Particulate fraction (PF) = 0.79 (dim. less)
SPM (= concentration of suspended particulate matter) = 4 mg/l
Theoretical surface water retention time (T_{SW}) = 8.3 days
Theoretical deep-water retention time (T_{DW}) = 8.4 days
Wave base, WB = $D_{\text{TA}} = 14.14 \text{ m}$

The concentration of nutrients in coastal areas can be influenced by emissions from different sources, e.g., from point-source emissions like fish farms and water purification plants. Since the theoretical surface water retention time of Baltic Sea coastal areas in the size range between 1 and 100 km^2 is often 2–10 days [63], it is evidently difficult to maintain high concentrations of nutrients from single additions, and, hence, difficult to use fertilization as an effective remedial measure for radionuclide concentrations in water and biota. Increasing the concentration of nitrogen or phosphorus also means probable changes in water clarity, gross sedimentation, and the amount of suspended particulate matter [81]. If realistic assumptions about possible changes in nutrient concentrations are made, the results given by [19] indicated that coastal water fertilization is not a relevant remedial measure. Firstly, because it would not be very efficient to reduce radionuclides in water, sediments and fish, secondly because it would not be acceptable for ecological reasons to fertilize coastal areas, since many European coastal areas (including the Baltic Sea) are already severely eutrophicated (see [70]), and thirdly because it would be very costly to try to maintain the higher nutrient concentrations.

In the following, results from a dredging scenario will be given, since this is one of the few feasible remaining options (see [1, 84] or Section [Remedial Measures for Aquatic Pathways](#) for compilations of remedial measures for aquatic systems).

Figure 28 gives results (see [20]) from a test in which 75% of the radionuclide in the active A-sediments was hypothetically removed month 7 (2 months after the simulated fallout), or months 9 or 12; and month 12 when also 75% of the radionuclides in the active A-sediments in the river were removed. Note that in practice, it is not realistic to remove as much as 75% of the inventory in A-sediments in large coastal areas, and it is not realistic to assume that this can be done shortly after the fallout. Thus, these simulations are meant to give indications as to whether dredging in coastal areas is advisable, and also to provide realistic expectations for such endeavors. The results are given for the concentration

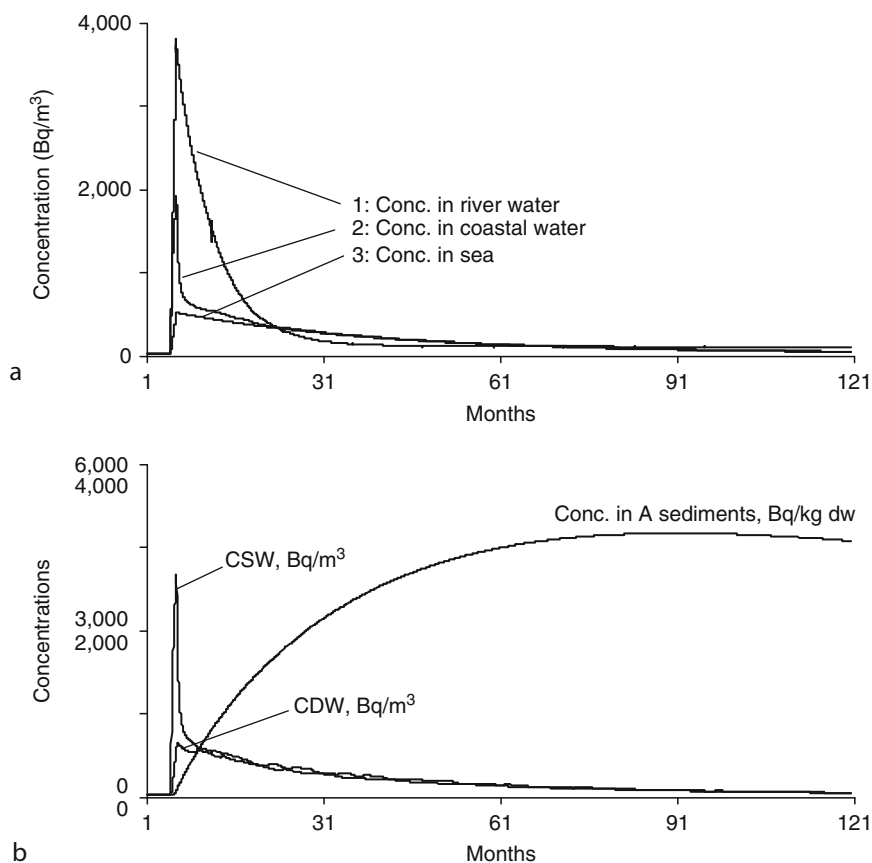


Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 24

A ranking of the fluxes to, from, and within the given Baltic Sea coastal area under default conditions (From [20])

in coastal water (Fig. 28a) and for the inventory of the radionuclide in A-sediments in (Fig. 28b). From these results, it should be noted that this would probably be a costly and inefficient remedy. Just like fertilization, it cannot be generally recommended in coastal areas. This means that actions like fish bans and domestic food treatment, e.g., salting of fish before consumption (see [1]), are likely to be more efficient than measures like fertilization and dredging (see Section Remedial Measures for Aquatic Pathways).

Discussion and Conclusions It is often stated that the quality of science is related to the possibilities for making meaningful predictions (see [21]). Models for hourly and daily predictions at different sites in coastal areas driven by climatological data on prevailing winds, precipitation, or temperatures may provide good descriptive power (if the necessary driving variables are at hand, which is not generally the case) but poor predictive power since it is difficult to predict winds and temperatures more than a few days ahead. The key



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 25

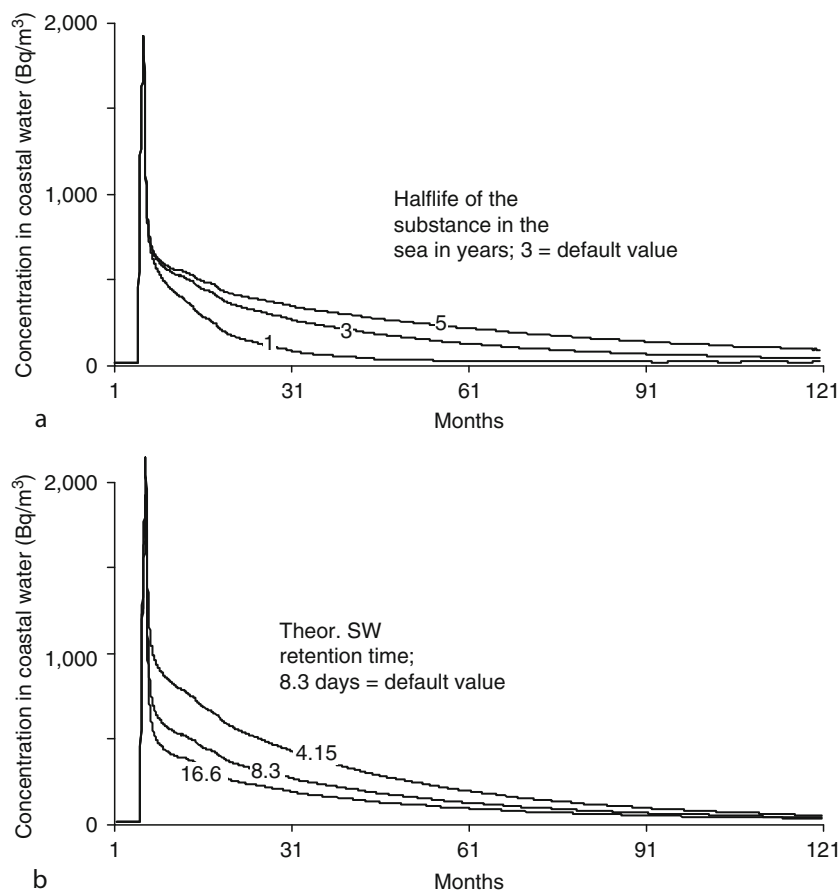
(a) A comparison between modelled radiocesium concentrations (Bq/m³) in river water, coastal water, and sea water (outside the given coastal area) under default conditions. (b) A comparison between modelled radiocesium concentrations (Bq/m³) in surface water, deep water (scale 0–4,000), and accumulation area sediments (A-sediments; scale 0–6,000 on the y-axis) in the default coastal area in the Baltic Sea (From [20])

to obtain predictive power lies in the structuring and equations used for the models, and models based on monthly predictions using ordinary differential equations are likely to be more optimal in contexts where practical applicability and high predictive power are requested.

The coastal model discussed here is based on the same model structuring as the lake and river models, which have shown to yield very good predictive power see (Figs. 12 and 17).

The method to define the coastal area and set the borderlines using the “topographical bottleneck”

approach (where the exposure attains minimum values) is very useful and a necessary prerequisite for the determination of three fundamental process variables for mass-balance models for coastal archipelagos and bays, [1] the theoretical surface water retention time (T_{SW}); [2] the theoretical deep-water retention time (T_{DW}); and [3] the bottom dynamic conditions (i.e., to define where resuspension from erosion- and transportation areas, ET, are likely to appear). This means that T_{SW} , T_{DW} , and ET can be estimated easily for such areas from standard bathymetric charts using GIS-methods. To estimate these variables using



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 26

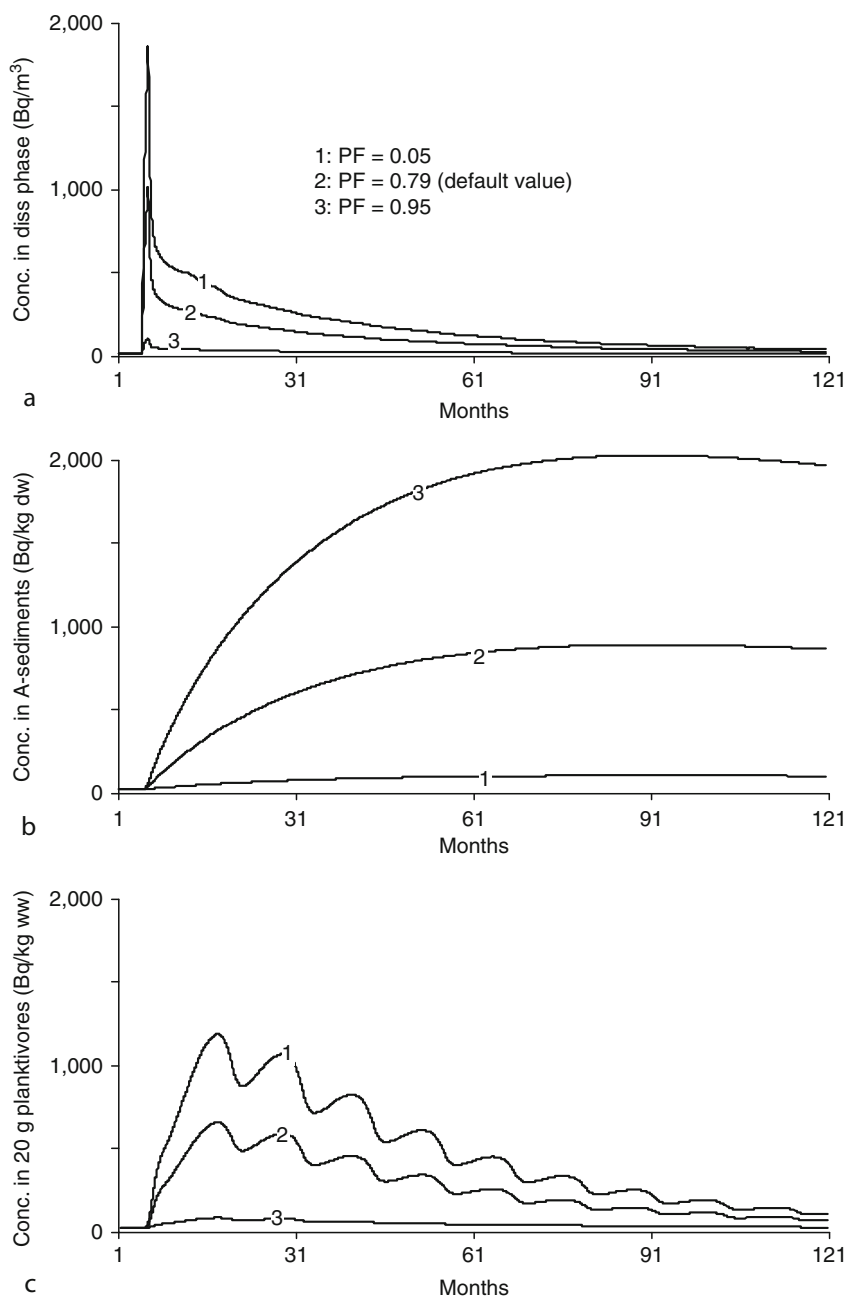
(a) Sensitivity analyses illustrating the role of different values for the ecological half-lives (1, 3, and 5 years) of the radionuclide in sea water outside the given coastal area for the concentration in coastal water. (b) Sensitivity analyses illustrating the role of different values for the theoretical surface water retention time (4.15, 8.3, and 16.6 days) for the concentration in the default coastal area in the Baltic Sea (From [20])

traditional methods (such as dye tracers, continuous monitoring, and/or field measurements) is very time consuming and expensive (see [19]).

The presented results have indicated that the predictions of radionuclide concentrations in water and fish largely depend on two factors, (a) the concentration in the sea outside the given coastal area and/or adjacent coastal areas and (b) the ecological half-life of the radionuclide in the sea. Uncertainties in these factors may dominate all other uncertainties, e.g., regarding the surface water retention time, the settling velocity of the particulate fraction of the radionuclide, and the distribution coefficient regulating the fluxes in

dissolved and particulate phases. This means that the conditions in the sea are of paramount importance for the conditions in the coastal area, even for relatively enclosed coastal areas.

Therefore, for any given coastal area one can produce a simplified version of the coastal model in which the relatively unimportant processes, factors, and sub-models are omitted and/or simplified. The river sub-model and the fish sub-model can be added to the coastal model for all large and open estuaries, and criteria for this have been discussed by [19]. The model can be used for this purpose for any coastal area. If the river sub-model and the

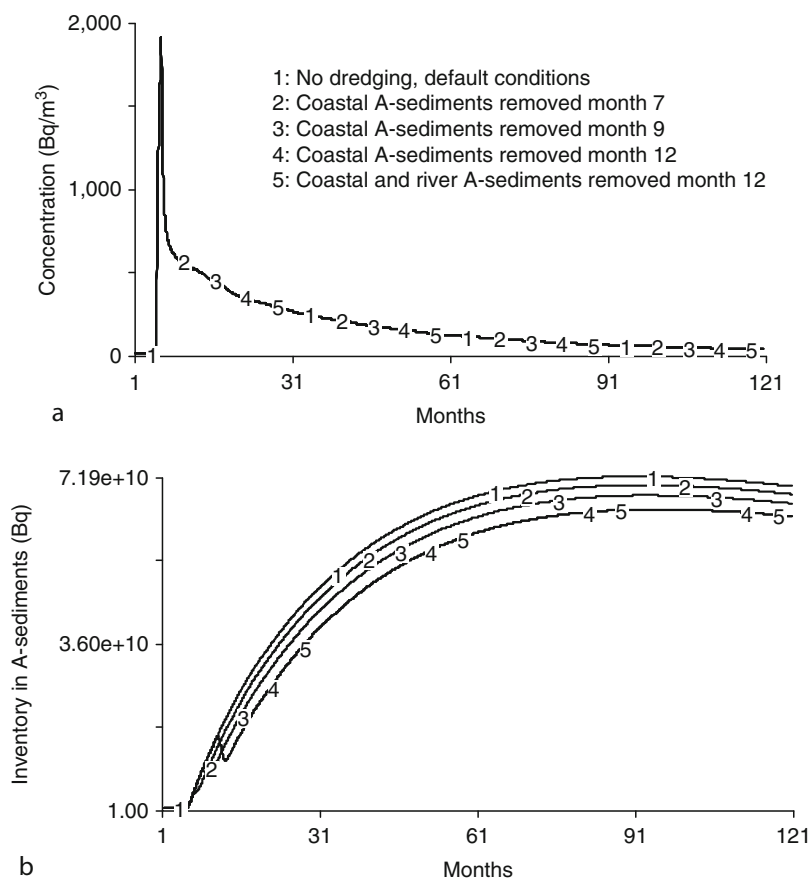


Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 27

Calculated concentrations in **(a)** dissolved phase in the coastal water, **(b)** in A-sediments, and **(c)** in 20 g planktivorous fish when the particulate fraction (PF) is set to 0.05 (curve 1), 0.79 (curve 2), and 0.95 (curve 3) in the default coastal area in the Baltic Sea (From [20])

fish sub-model are added to the basic coastal mass-balance model, this implies that the catchment area parameters should also be added to the panel of driving variables.

This coastal model may be regarded as a scientific tool for posing and testing working hypotheses on the relative roles of different processes in different coastal areas. Such information may be essential for getting



Nuclear Accidents, Chernobyl Fallout in Scandinavian Watersheds. Figure 28

Results of simulated dredging operations when 75% of the radionuclides in the coastal A-sediments have been removed (months 7, 9, and 12; curves 2, 3, 4, and 5) and 75% of the river A-sediments (month 12; curve 5) calculated results for (a) the concentration in coastal water and (b) for the inventory in coastal A-sediments in the default coastal area in the Baltic Sea (From [20])

realistic expectations of various remedial measures, such as coastal dredging discussed above.

Remedial Measures for Aquatic Pathways

Introduction and Aim

There are a number of possible intervention measures to reduce radioactive doses to the public via the water pathways. Countermeasures may be grouped into two categories: those aimed at reducing radionuclides in drinking water and those reducing doses to man from consumption of aquatic foodstuffs, principally fish. It has been shown [30, 85] that doses to man from

terrestrial foodstuffs are generally more significant than doses from drinking water and aquatic foodstuffs. Measures to reduce doses via drinking water may, however, be required in the short term (weeks) after a fallout when radionuclide concentrations in surface waters are relatively high. Measures to reduce doses via fish may be required over longer periods as a result of bioaccumulation of radioactivity through the aquatic food chain. In addition, it was found [85] that after the Chernobyl accident the public perceived that the contamination of water supplies was a much greater problem than it actually was. Thus, measures to protect water supplies were also used as a method of restoring

public confidence. It should be noted that surface waters are important transporters of radioactivity from contaminated zones to populations in less contaminated areas. The use of water supplies for drinking and irrigation generally results in small individual doses, but may lead to significant collective doses.

The purpose of this section is to discuss options available to decision makers in the event of radioactive contamination of surface waters and the results discussed here are mainly derived from a paper by [84]. The framework is the requirement for countermeasures in the context of a large-scale atmospheric release of radionuclides. The following intervention measures will be discussed:

- Drinking water distribution and treatment
- Direct and secondary contamination of surface waters
- Uptake by fish and aquatic foodstuffs from contaminated water
- Preparation of fish prior to consumption

Measures Related to Drinking Water Consumption

Temporary bans on drinking water consumption and provision of alternative supplies may be required immediately after a fallout. Since radionuclide concentrations in rivers tend to decline rapidly (half-life of decline of Cs-137 and Sr-90 in rivers during this early period is typically 2 weeks and of I-131 about 5–6 days), such bans are not likely to be in place for more than a few weeks after fallout.

Switching to less contaminated water supplies may be an accepted and effective countermeasure to radioactive contamination of drinking water. If the water supply system is sufficiently integrated, supplies may rapidly be switched to groundwater sources, or less contaminated surface water sources. The effectiveness of this approach depends on a quick assessment of the extent of the contamination of different water supplies.

Supplementary purification of drinking water in water treatment plants is another option. The mechanism for removal is generally by attachment to and settling of solid particles during flocculation, as well as removal during filtration. It has been found that activated charcoal was effective in removing I-131 and Ru-106 and zeolite in removing and radiostrontium [86]. These substances were found to

be effective for the first 3 months, after which they became saturated and their efficiency declined [86].

Regulation of flow of contaminated water through reservoirs is another viable action. After Chernobyl, surface gates on the Kiev Reservoir dam were opened to let clean water out of the reservoir so that surface runoff water from the contaminated areas could be captured by the reservoir. Because the relative importance of atmospheric fallout and secondary contamination processes to reservoir and lake contamination was not known, the countermeasure had the opposite effect to that intended. In fact, because of atmospheric fallout to the reservoir surface, deep waters in the reservoir were much less contaminated than surface ones.

Measures to Reduce Contamination of Surface Waters

Various measures could be used to reduce runoff of radionuclides attached to eroded soil particles (see, e.g., [87–89]). Typically, suspended solid concentrations in river and lake waters are lower than 50 mg/l and less than 50% of radiocesium and less than 10% of radiostrontium and radioiodine are expected to be in the particulate phase, limiting the potential effectiveness of this method. It should also be noted that the dissolved, rather than particulate, forms of these radionuclides are important in determining concentrations in drinking water and freshwater biota.

Dredging of canal-bed traps to intercept suspended particles in contaminated rivers was carried out after Chernobyl. These traps were found to be inefficient for two reasons: [1] flow rates were too high to trap small suspended particles carrying much of the radioactivity; [2] a significant proportion of the activity (and most of the “available” activity) was in dissolved forms and thus would not have been intercepted by the sediment traps.

Buffer strips adjacent to rivers can be used to trap eroded particles and increase sorption of radionuclides to soil. Buffer strips have been used to reduce runoff of fertilizers and pesticides from agricultural land, but are untested for radionuclides. Removal of radioactivity from water by “filtering” through reed beds is also possible, though probably not feasible on a large scale. Zeolite-containing dykes were constructed on smaller

rivers and streams around Chernobyl in order to intercept dissolved radionuclides. These were found to be ineffective [85]. In addition, the rivers and streams on which they were placed were later found to contribute only a few percent to the total activity load in the Pripyat–Dnieper system.

Measures to Reduce Uptake by Fish and Aquatic Foodstuffs

Bans on consumption of freshwater fish have been applied in the restricted zones of Chernobyl [90]. In some areas, selective bans on the more contaminated predatory fish have been applied. It seems that bans are often ignored by fishermen. Bans on the sale of freshwater fish have been applied in some areas of Norway [91]. Use of farmed fish has been suggested by [92] as an alternative source of freshwater fish in areas affected by fishing bans. Farmed fish, fed with uncontaminated food, do not significantly accumulate radionuclides [92, 93].

Lake liming was tested in 18 Swedish lakes [11]. Different types of lime were used: primary rock lime (PR) was used in ten lakes, sedimentary rock lime in six lakes, and so-called mixed lime (a lime also containing various trace elements and nutrients) in two lakes. The lime was spread over the lakes by different methods, such as helicopters, pontoon boats, and dosers, in the mouth of the tributaries. The amounts of lime supplied were as one or several applications, either on ice or directly into the water. As a result of the liming, the mean pH of the lakes increased and also other parameters that are directly linked to the addition of lime, e.g., hardness and alkalinity. A certain decrease in the color of the lakes was also noted. The results of the experiments showed that liming had no significant effect on uptake of Cs-137 in fish in comparison with control lakes. Although uptake of Sr-90 was not studied in these experiments, it is expected that increased Ca concentration in lakes may reduce S-concentration in fish.

Wetland liming was carried out in the catchment areas of 17 lakes, of which eight were limed with primary rock lime, seven with sedimentary lime, and two with “mixed lime.” The amount of lime added followed the recommendations by the Swedish Environmental Protection Agency, which implies 0.3 t/ha totally over the catchment area, 15 t/ha on wetlands where the

wetlands make up 2% of the area, and 3 t/ha on wetlands where the wetlands make up 10% of the area (see [19]). The supply was done using helicopters and a single application. In theory, wetland liming has several advantages in comparison with lake liming:

- A considerably prolonged durability.
- A possible “liming shock” in the lake is minimized.
- Improved conditions for animals and plants in streams and rivers; reduced transport of several metals (e.g., Fe and Al) into the lake from the catchment area.

Full-scale drainage area liming was done in two areas using dolomite. Since the dolomite is less soluble than ordinary limestone, the durability should be further prolonged.

Although the mean pH in the limed lakes increased from 6.0 to 6.7, and also other variables, which are directly linked to the liming remedies (e.g., water hardness and alkalinity) showed increased long-term mean values, no effect of the approach was observed in Cs-concentrations in fish.

Addition of potassium to lakes: It is known that the concentration factor of radiocesium in fish is inversely related to the potassium content of the surrounding water (see [94, 95]). After Chernobyl, potassium was added to 13 lakes in Sweden either as potash or as an additive in the mixed lime [11]. In most of the treated lakes, mean potassium concentrations were increased from less than 0.4 mg/l to about 0.8 mg/l. The results of the potash treatment were somewhat inconclusive, with a small reduction in concentrations in perch fry being observed during the 2-year experiment. It was found that in lakes with short water retention times, it was difficult to maintain high levels of potassium in the lake. It is likely that potassium treatment is only feasible in lakes with very long water residence times, allowing increased potassium concentrations to be maintained. In an experiment on Lake Svyatoye in Belarus, Smith et al. (1999) added potassium chloride fertilizer onto the frozen lake surface. This resulted in an increase of the lake-water potassium concentration from 1 to 10 mg/l. Because of the long water residence time of this lake (it has no outflow), potassium concentration was still 8 mg/l 1 year after application. Initial results showed a 50% reduction in Cs-137 concentration in fish during the first year after the experiment [92, 96].

Manipulation of the aquatic food web – intensive fishing and fertilization: Intensive fishing was carried out in four lakes in Sweden [11] and as a complementary measure in three other lakes. This resulted in a reduction of the fish population by about 5–10 kg per hectare. The species reduced were mainly pike, perch, and roach. No effect of intensive fishing on Cs-137 concentrations in fish was observed. Fertilization was carried out in two Swedish lakes using “Osmocoat” (5% P and 15% N). The concentrations of total-P generally showed no significant changes in the long-term mean value: It appears that the fertilization remedy was not conducted sufficiently effectively. No effect was observed on Cs-137 activity concentrations in fish.

Preparation of Fish Prior to Consumption

Different methods of food preparation may affect the quantity of radionuclides in consumed food. Ryabov [90] suggested bans on consumption of smoked and dried fish since these processes increase concentrations of radionuclides (per unit of weight consumed). Other preparation processes may reduce radioactivity in fish.

Although radiocesium is relatively evenly distributed throughout the fish, radiostrontium is strongly concentrated in the bony parts of fish. An effective measure to reduce consumption of radiostrontium is to remove the bony parts of fish prior to cooking.

It has been known for a long time that salting may eliminate Cs-137 from meat. After 2–4 days immersion in concentrated salt solution, the Cs-concentration in a 500 g piece of meat may be reduced by 60–70% [97]. There are, however, at least three severe drawbacks with the salting method:

1. The amount of vitamins (e.g., B6, B12) would decrease significantly with the time of exposure; after 2–4 days most of these vitamins would be lost.
2. The levels of Na (and K) would increase and this would influence the taste of the meat; a large intake of Na and K may also give rise to negative health effects in humans.
3. The color and texture of the meat would change, and most people would probably agree that the grayer and harder meat obtained after 2–4 days in a salt solution would be less appetizing than normal, fresh meat.

Tests were carried out [97] to try to speed up the salting procedure from 2 to 4 days to 2–4 h. Different types of salts were tested, as well as different types of meat, and different forms and sizes of meat. Two different approaches to speed up the salting procedure were also tested: rotating the meat in the salt solution (called centrifugation) or to rotate the salt solution around the meat (called rotation). Both rotation and centrifugation in salted water may reduce Cs-137 in meat (moose, deer, and fish) by 60–70% in 2–3 h. The first round (of centrifugation or rotation) was most efficient in reducing Cs-137: about 30–50% of the Cs-137 may be reduced in the first round (depending on the salt concentration and the type of meat), with about 10–15% reduction in subsequent rounds. There is no significant change in vitamin B6 if the treatment lasts for 2–3 h. There is also no significant change in the texture of the meat, except an increase in the Na concentration. The best results in terms of Cs-137 reduction were obtained for meat that had been frozen: frozen meat showed a 20–40% decrease in Cs-137 compared to unfrozen conditions.

The rotation and centrifugation procedure takes 2–3 h and it requires access to a household machine that could either centrifugate the meat (about 1,000 rotation/min.) or rotate the salt solution. Salting, speeded up by rotation, may be a practical method in areas with heavily contaminated meat, especially for larger kitchens (schools, military, restaurants, etc.). It may also be used in homes, because there may be people who, from primarily a psychological point of view, would like to feel that they are able to do something themselves to reduce their intake of radioactive cesium.

Conclusions

The key to successful implementation of countermeasures is the provision of accurate information to the public. Measures can be counterproductive in increasing the stress people feel in a situation of radioactive contamination. If not accompanied by adequate information, the application of countermeasures may make people overestimate the danger of their exposure to radiation. People may feel that if countermeasures are necessary, the potential exposure must be “dangerous,” even when actual radiation doses present only a small

risk. Conversely, countermeasures may also be implemented to increase people's fears of radioactivity. Thus, after Chernobyl, measures were implemented to reduce radioactivity in the Kiev Reservoir even though doses via this pathway to the people of Kiev were relatively low.

This section has discussed a number of measures to reduce doses of radionuclides via the freshwater pathway. These range from simple and well-tested measures (e.g., bans on drinking water consumption) to more complex measures such as lake liming or potassium addition. The most effective and viable measures to reduce radioactivity in drinking water are those that operate at the water treatment and distribution stage. Intervention measures to reduce concentrations of radioactivity in rivers and reservoirs are expected to be less efficient at reducing doses via the drinking water pathway for two primary reasons:

- It is generally only possible to apply intervention measures to reduce concentrations of radioactivity in rivers and reservoirs in the medium term (months) after an accident since they require significant engineering works. Radioactivity concentrations in drinking water tend largely to be a short-term problem – the concentrations in river water decline rapidly during the first days to few weeks after a fallout. Thus, intervention needs to be implemented rapidly and therefore based on accurate monitoring data and modelling.
- Intervention to reduce radioactivity concentrations in rivers and reservoirs necessarily requires manipulation of much larger volumes of water than intervention at the water treatment and supply stage.

Direct measures to reduce uptake of radioactivity by fish, and food preparation measures, are most likely to be effective at reducing doses via consumption of fish. Deboning fish contaminated by radiostrontium is probably the most effective food preparation measure, but salting and freezing can also reduce radiocesium concentrations. Bans on consumption of fish can be effective, but there are few viable measures to reduce radioactivity in fish prior to the preparation stage. Lake liming and biomanipulation were ineffective for radiocesium, though addition of potassium to lake waters appears promising in lakes with long water

retention times. Lake liming may be effective in reducing radiostrontium in fish, though this has not, to our knowledge, been tested.

Future Directions

It is essential in radioecology to predict the outcome of remedial measures designed to lower doses of radionuclides to man and ecosystems and evidently also to predict the consequences of future possible accidents. The predictive power of quantitative models in radioecology need to be high and to achieve that goal, it is important to initiate and run international comparative projects. That means that there is a need to continue with projects such as the famous VAMP project and the BIOMOVs project based on large amounts of reliable empirical data and strict model comparisons.

Bibliography

Primary Literature

1. IAEA (2000) Modelling of the transfer of radiocesium from deposition to lake ecosystems. Report of the VAMP Aquatic Working Group. International Atomic Energy Agency, Vienna, IAEA-TECDOC-1143, p 343
2. Sansone U, Voitsekhovitch OV (1996) Modelling and study of the mechanisms of the transfer of radionuclides from the terrestrial ecosystem to and in water bodies around Chernobyl. EUR 16529 EN. European Commission, Luxembourg
3. Strand P, Howard B, Averin V (eds) (1996) Transfer of radionuclides to animals, their comparative importance under different agricultural ecosystems, and appropriate countermeasures. EUR 16539 EN. European Commission, Luxembourg
4. Moberg L (ed) (1991) The Chernobyl fallout in Sweden. The Swedish Radiation Protection Institute, Stockholm, p 633
5. Onishi Y, Voitsekhovich O, Zheleznyak MJ (2007) Chernobyl – what have we learned? Springer, Heidelberg, p 289
6. Monte L, Brittain JE, Håkanson L, Gallego E (1999) MOIRA models and methodologies for assessing the effectiveness of countermeasures in complex aquatic systems contaminated by radionuclides. ENEA, RT/AMP, Rome, p 150
7. Monte L, Brittain JE, Håkanson L, Heling R, Smith JT, Zheleznyak M (2003) Review and assessment of models used to predict the fate of radionuclides in lakes. J Environ Radioact 69:177–205
8. Monte L, Brittain JE, Håkanson L, Smith JT, van der Perk M (2003) Review and assessment of models for predicting the migration of radionuclides from catchments. J Environ Radioact 75:83–103
9. Monte L, Boyer P, Brittain JE, Håkanson L, Lepicard S, Smith JT (2005) Review and assessment of models for predicting the

- migration of radionuclides through rivers. *J Environ Radioact* 79:273–296
10. Andersson T, Håkanson L, Kvarnäs H, Nilsson Å (1991) Remedial measures against high levels of radioactive cesium in Swedish lake fish (in Swedish). SSI Report 91-07, p 114
 11. Håkanson L, Andersson T (1992) Remedial measures against radioactive cesium in Swedish lake fish after Chernobyl. *Aquatic Sci* 54:141–164
 12. Abrahamsson O, Håkanson L (1997) Presentation and analysis of a model simulating the response of a potash treatment of a lake. *J Environ Radioact* 37:287–306
 13. Andersson E (1989) Incorporation of Cs-137 into fishes and other organisms. In: Feldt W (ed) *The radioecology of natural and artificial radionuclides*. Verlag TUV, Köln, pp 312–317
 14. Dahlgaard H (ed) (1994) *Nordic radioecology, the transfer of radionuclides through nordic ecosystems to man*. Elsevier Science, Amsterdam, p 483
 15. Monte L, Håkanson L, Brittain J (1997) Prototype models for the MOIRA computerised system. ENEA, ISSN/1120-5555, Roma, p 90
 16. Monte L (1996) Collective models in environmental science. *Sci Total Environ* 192:41–47
 17. Håkanson L, Peters RH (1995) *Predictive limnology. Methods for predictive modelling*. SPB Academic Publishing, Amsterdam, p 464
 18. Monte L (1995) A simple formula to predict approximate initial contamination of lake water following a pulse deposition of radionuclide. *Health Phys* 68(3):397–400
 19. Håkanson L (2000) Modelling radiocesium in lakes and coastal areas - new approaches for ecosystem modellers. A textbook with Internet support. Kluwer, Dordrecht, p 215
 20. Håkanson L (2005) A new general dynamic model predicting radionuclide concentrations and fluxes in coastal areas from readily accessible driving variables. *J Environ Radioact* 78: 217–245
 21. Peters RH (1991) *A critique for ecology*. Cambridge University Press, Cambridge, p 366
 22. Knoechel R, Campbell CE (1988) Physical, chemical, watershed and plankton characteristics of lakes on the Avalon Peninsula, Newfoundland, Canada: a multivariate analysis of interrelationships. *Verh Internat Verein Limnol* 23:282–296
 23. Rochelle B, Liff C, Campbell W, Cassell D, Church R, Nusz R (1989) Regional relationships between geomorphic/hydrologic parameters and surface water chemistry relative to acidic deposition. *J Hydrol* 112:103–120
 24. Santschi PH, Honeyman BD (1991) Radioisotopes as tracers for the interactions between trace elements, colloids and particles in natural waters. In: Vernet JP (ed) *Heavy metals in the environment*. Elsevier, Amsterdam
 25. Riise G, Björnstad HE, Oughton DH, Salbu B (1990) A study on radionuclide associations with soil components using sequential extraction procedure. *J Radioanal Nucl Chem* 142:531–538
 26. Erel Y, Stolper EM (1993) Modeling of rare-earth element partitioning between particles and solution in aquatic environments. *Geochim Cosmochim Acta* 57:513–518
 27. Gustafsson Ö, Gschwend PM (1997) Aquatic colloids: Concepts, definitions and current challenges. *Limnol Oceanogr* 42:519–528
 28. Comans RNJ (ed) (1998) *Modelling fluxes and bioavailability of radiocesium and radiostrontium in freshwaters*. Progress report from the ECOPRAQ project. Netherlands Energy Research Foundation
 29. Comans RNJ, Hockley DE (1992) Kinetics of cesium sorption on illite. *Geochim Cosmochim Acta* 56:1157–1164
 30. Stone D, Smith JT, Jackson D, Ibbotson AT (1997) Scoping study on the relative importance of freshwater dose pathways following a major nuclear accident. Report to IMC/Magnox Electric 1997
 31. Håkanson L (1991) Ecometric and dynamic modelling – exemplified by cesium in lakes after Chernobyl. Springer, Berlin, p 158
 32. Lindström M (2000) Predictive modelling of heavy metals in urban lakes. Doctoral thesis, Uppsala University, Sweden
 33. Håkanson L, Sazykina TG, Kryshev II (2002) A general approach to transform a lake model for one radionuclide (radiocesium) to another (radiostrontium) and critical model tests using data for four Ural lakes contaminated by the fallout from the Kyshtym accident in 1957. *J Environ Radioact* 60:319–350
 34. Förstner U, Wittmann GTW (1981) *Metal pollution in the aquatic environment*. Springer, Berlin, p 486
 35. Salomons W, Förstner U (1984) *Metals in the hydrocycle*. Springer, Heidelberg, p 349
 36. Kirkby HJ (ed) (1978) *Hillslope hydrology*. Wiley, Chichester, p 389
 37. Hillel D (1982) *Introduction to soil physics*. Academic, New York, p 364
 38. Harper DM, Stewart WDP (1987) The effects of land use upon water chemistry, particularly nutrient enrichment, in shallow lowland lakes: comparative studies in three lochs in Scotland. *Hydrobiologia* 148:211–229
 39. Schofield NJ, Ruprecht JK (1989) Regional analysis of stream salinisation in southwest Western Australia. *J Hydrol* 112: 19–39
 40. Alberts JJ, Tilly LJ, Vigerstad TJ (1979) Seasonal cycling of cesium-137 in a reservoir. *Science* 203:649–651
 41. Håkanson L (2004) A new generic sub-model for radionuclide fixation in large catchments from continuous and single pulse fallouts, as used in a river model. *J Environ Radioact* 77: 247–273
 42. Black VS (1957) Excretion and osmoregulation. In: Brown ME (ed) *The physiology of fishes*, vol 1. Academic, New York, pp 163–205
 43. Fleishman DG (1963) Accumulation of artificial radionuclides in freshwater fish. In: Klechkovskii VM, Polikarpov GG, Aleksakhin RM (eds) *Radioecology*. Wiley, New York, pp 347–370
 44. Carlsson S (1978) A model for the turnover of Cs-137 and potassium in pike (*Esox Lucius*). *Health Phys* 35:549–554
 45. Madruga MJ, Cremers A (1997) On the differential binding mechanisms of radiostrontium and radiocesium in sediments.

- In: Desmet G et al (eds) *Freshwater and estuarine radioecology*. Elsevier, Amsterdam, pp 207–216
46. Konitzer K, Meili M (1997) Redistribution of sedimentary Cs-137 in small Swedish lakes after the Chernobyl fallout 1986. In: Desmet G et al (eds) *Freshwater and estuarine radioecology*. Elsevier, Amsterdam, pp 167–172
 47. Konoplev A, Bulgakov A, Hilton J, Comans R, Popov V (1997) Long-term kinetics of radiocesium fixation by soils. In: Desmet G et al (eds) *Freshwater and estuarine radioecology*. Elsevier, Amsterdam, pp 173–182
 48. Håkanson L, Jansson M (1983) *Principles of lake sedimentology*. Springer, Berlin, p 316
 49. Håkanson L, Boulion VV (2002) Empirical and dynamical models to predict the cover, biomass and production of macrophytes in lakes. *Ecol Model* 151:213–243
 50. Kranck K (1973) Flocculation of suspended sediment in the sea. *Nature* 246:348–350
 51. Kranck K (1979) Particle matter grain-size characteristics and flocculation in a partially mixed estuary. *Sedimentology* 28:107–114
 52. Lick W, Lick J, Ziegler CK (1992) Flocculation and its effect on the vertical transport of fine-grained sediments. *Hydrobiologia* 235(236):1–16
 53. Sly PG (1978) *Sedimentary Processes in Lakes*. In: Lerman A (ed) *Lakes: chemistry, geology, physics*. Springer, Berlin, pp 65–89
 54. Thomas RL (1972) The distribution of mercury in the sediments of Lake Ontario. *Can J Earth Sci* 9:636–651
 55. Thomas RL, Kemp ALW, Lewis CFM (1972) Distribution, composition and characteristics of the surficial sediments of Lake Ontario. *J Sediment Petrol* 42:66–84
 56. Håkanson L (1977) The influence of wind, fetch, and water depth on the distribution of sediments in Lake Vänern, Sweden. *Can J Earth Sci* 14:397–412
 57. Bloesch J, Burns NM (1980) A critical review of sedimentation trap technique. *Schweiz Z Hydrol* 42:15–55
 58. Bloesch J, Uehlinger U (1986) Horizontal sedimentation differences in a eutrophic Swiss lake. *Limnol Oceanogr* 31: 1094–1109
 59. Einstein HA (1950) The bed-load function for sediment transportation in open-channel flows, Soil Conservation Service. Technical Bulletin, no. 1026. US Department of Agriculture, Washington, DC
 60. Allen JRL (1970) *Physical processes of sedimentation*. Allen & Unwin, London, p 248
 61. Wetzel RG (1983) *Limnology*. Saunders College Publishing, Philadelphia, p 767
 62. Walker CH, Hopkin SP, Sibly RM, Peakall DB (2001) *Principles of ecotoxicology*, 2nd edn. Taylor & Francis, London, p 309
 63. Håkanson L (1999) *Water pollution – methods and criteria to rank, model and remediate chemical threats to aquatic ecosystems*. Backhuys Publishers, Leiden, p 277
 64. Rosenberg R (ed) (1984) *Biological evaluation of shallow Swedish coastal areas (in Swedish, Biologisk värdering av grunda svenska havsområden)*. Swedish EPA, report 1911, p 384
 65. Mann KM (1992) *Ecology of coastal waters – a systems approach*. Blackwell Scientific Publications, Oxford, p 322
 66. Lundin LC (1999) *Water in society. Sustainable water management in the Baltic Sea basin*, The Baltic University Programme, Uppsala University, Sweden, p 244
 67. Lundin LC (2000) *The waterscape. Sustainable water management in the Baltic Sea basin*, The Baltic University Programme, Uppsala University, Sweden, p 207
 68. Lundin LC (2000) *River basin management. Sustainable water management in the Baltic Sea basin*, The Baltic University Programme, Uppsala University, Sweden, p 244
 69. Meeuwig JJ, Kauppila P, Pitkänen H (2000) Predicting coastal eutrophication in the Baltic: a limnological approach. *Can J Fish Aquat Sci* 57:844–855
 70. Aertbjerg G (ed) (2001) *Eutrophication in Europe's coastal waters*. European Environment Agency, Topic report 7/2002, Copenhagen, p 86
 71. BIOMOVs (1991) *Dynamics within lake ecosystems. Scenario A5*. Sundblad B (ed) Technical report 12, September 1991, Swedish Radiation Protection Institute
 72. Davis P (1999) BIOMOVs II. Special Issue. *J Environ Radioact* 42:1–304
 73. Davis PA, Avadhanula MR, Cancio D, Carboneras P, Coughtrey P, Johansson G, Little RH, Smith GM, Watkins BM (1999) BIOMOVs II: An international test of the performance of environmental transfer models. *J Environ Radioact* 42: 117–130
 74. Smith JT, Konoplev AV, Bulgakov AA, Comans RNJ, Cross MA, Kaminski S, Khristuk B, Klemm E, de Konig A, Kudelsky AV, Laptev G, Madruga MJ, Voitsekhovitch OV, Zibild G (2002) AQUASCOPE Technical Deliverable. Simplified models for predicting ^{89}Sr , ^{90}Sr , ^{134}Cs , ^{137}Cs , ^{131}I in water and fish of rivers, lakes and reservoirs. CEH Centre for Ecology and Hydrology, Natural Environmental Research Council
 75. Zheleznyak M, Shepeleva V, Sizonenko V, Mezheva I (1997) Simulation of countermeasures to diminish radionuclide fluxes from the Chernobyl zone via aquatic pathways. *Radiat Prot Dosim* 73:181–186
 76. Monte L (1997) A collective model for predicting the long-term behaviour of radionuclides in rivers. *Sci Total Environ* 201:17–29
 77. Monte L (1998) Predicting the long term behaviour of ^{90}Sr in lacustrine systems by a collective model. *Ecol Modell* 106:141–159
 78. Smith JT, Wright SM, Cross MA, Monte L, Kudelsky A, Saxen R, Vakulovsky SM, Timms DN (2004) Global analysis of the riverine transport of ^{90}Sr and ^{137}Cs . *Environ Sci Technol* 38:850–857
 79. Monte L, Van der Steen J, Bergström U, Gallego Díaz E, Håkanson L, Brittain J (2000) The project MOIRA A model-based computerised system for management support to identify optimal remedial strategies for restoring radionuclide contaminated aquatic ecosystems and drainage areas. Final report. ENEA. ISSN/1120-5555, Rome, p 89

80. Appleby PG, Oldfield F (1979) Letter on the history of lead pollution in Lake Michigan. *Environ Sci Technol* 13:263–273
81. Håkanson L (2006) Suspended particulate matter in lakes, rivers and marine systems. The Blackburn Press, New Jersey, p 319
82. Pilesjö P, Persson J, Håkanson L (1991) Digital bathymetric information for calculations of coast morphometrical parameters and surface water retention time (in Swedish with English summary). Swedish Environmental Protection Agency, Report 3916, Stockholm, p 76
83. Håkanson L, Eklund JM (2007) A dynamic mass-balance model for phosphorus fluxes and concentrations in coastal areas. *Ecol Res* 22:296–320
84. Smith JT, Voitsekrovitch OV, Håkanson L, Hilton J (2001) A critical review of measures to reduce radioactive doses from drinking water and consumption of freshwater food-stuffs. *J Environ Radioact* 56:11–32
85. Voitsekrovitch OV, Nasvit O, Los'y Y, Berkovski V (1997) Present thoughts on aquatic countermeasures applied to regions of the Dnieper river catchment contaminated by the 1986 Chernobyl accident. In: Desmet G, Blust R, Comans RNJ, Fernandez J, Hilton J, de Bettencourt A (eds) *Freshwater and estuarine radioecology*. Elsevier, Amsterdam, pp 75–86
86. Tsarik N (1993) 1993, Supplying water and treating sewage in Kiev after the Chernobyl accident. *J Am Water Works Assn* 85(1):42–45
87. Schwab GO, Frevert RK, Edminster TW, Barnes KK (1966) *Soil and water conservation engineering*. Wiley, New York
88. Morgan RPC (1986) *Soil erosion and conservation*. Longman, Harlow
89. Hudson NW (1981) *Soil conservation*. Batsford, London
90. Ryabov IN (1992) Analysis of countermeasures to prevent intake of radionuclides via consumption of fish from the region affected by the Chernobyl accident. In: *Proceedings of international seminar on intervention levels and countermeasures for nuclear accidents*, EUR 14469, CEC, Luxembourg, pp 379–390
91. Brittain JE, Storruste A, Larsen E (1991) Radiocaesium in Brown Trout (*Salmo trutta*) from a subalpine lake ecosystem after the Chernobyl reactor accident. *J Environ Radioact* 14:181–191
92. Smith JT, Kudelsky AV, Ryabov IN, Hadderingh RH (1999) Aquatic countermeasure against radiocaesium uptake by the ecosystem (AQUACURE) second interim report. European Commission, 1999
93. Smith JT, Leonard DRP, Hilton J, Appleby PG (1997) Towards a generalized model for the primary and secondary contamination of lakes by Chernobyl-driven radiocesium. *Health Phys* 72:880–892
94. Camplin WC, Leonard DRP, Tipple JR, Duckett L (1989) Radioactivity in freshwater systems in Cumbria (UK) following the Chernobyl accident MAFF fisheries research data report no. 18, MAFF, Lowestoft
95. Fleishman DG, Nikiforov VA, Saulus AA, Komov VT (1973) Cs-137 in fish of some lakes and rivers of the Bryansk region and North-West Russia in 1990–1992. *J Environ Radioact* 24:145–158
96. Blaylock BG (1982) Radionuclide data bases available for bioaccumulation factors for freshwater biota. *Nucl Safety* 23:427–438
97. Smith JT, Leonard DRP, Hilton J, Appleby PG (1997) Towards a generalized model for the primary and secondary contamination of lakes by Chernobyl-driven radiocesium. *Health Phys* 72:880–892
98. Wallström A, Håkanson L (1991) A household method to reduce Cs-137 in meat. Swedish Health and Food Institute, Stockholm (in Swedish)
99. Andersson P (2006) Miljöövervakning före och efter Tjernobylolyckan. *Strålskyddsnytt* 1:37–38
100. Johansson KJ (2006) Cesium-137 i svamp, bär och vilt. *Strålskyddsnytt* 1:25–31
101. Håkanson L (1997) Modelling of radiocesium in lakes - on predictive power and lessons for the future. In: Desmet G et al (eds) *Freshwater and estuarine radioecology*. Elsevier, Amsterdam, pp 3–45

Books and Reviews

Salomons W, Förstner U (1984) *Metals in the hydrocycle*. Springer, Heidelberg, p 349

Nuclear Energy, Introduction

NICHOLAS TSOUFANIDIS

Department of Nuclear Energy, Missouri University of Science and Technology, Rolla, MO, USA

In terms of technical progress of the human species/society, the second half of the twentieth century is marked by two developments: the computer and nuclear energy. And the two are related since progress in the development and applications of nuclear energy owes a lot to the power of computations made possible by the digital computer.

The whole twentieth century is marked by the ever-increasing use of electricity. The century started with a tiny amount of electricity use and ended with ~30% of the total energy consumed to be in the form of electricity.

Nuclear energy, unfortunately, was released in the world as a weapon. But, fortunately, after the initial shock of Hiroshima and Nagasaki, nuclear energy turned out to be of great benefit to society. First, in a “Faustian bargain,” of sorts, the existence of nuclear weapons resulted in having a *cold war* between the two

superpowers of the time (Soviet Union and the USA) instead of a *hot one*. Second, nuclear energy proved to be an excellent method of generating electricity and also provided the means for numerous applications in industry, science, education, and probably most of all medicine. Not even the fiercest critics of nuclear energy deny the benefits of nuclear medicine in correct diagnosis and therapy and, therefore, prolongation of life for millions of people all over the world.

Although, as mentioned above, nuclear energy touches today beneficially many aspects of our lives, the most prominent application is the generation of electricity by using nuclear fission reactors to generate heat that is used to produce steam (or hot gas) that runs a turbo generator and produces electricity. There are many advantages in using nuclear energy for the generation of electricity. Primary among them is the absence of green house gases and other pollutants such as sulfur oxides, nitrogen oxides, heavy metals (mercury etc.).

Probably, the most important difference between nuclear and fossil or any renewable fuels is the concentrated energy release of the former. The fission of one U nucleus releases $\sim 200 \times 10^6$ eV; the burning of one atom/molecule in a chemical reaction releases a few eV; a ratio of many million! A single pellet of nuclear fuel, about the size of a fingertip, contains as much energy as 1,700 ft³ of natural gas, or 1,780 lb of coal, or 149 gal of oil. As a result of this concentrated energy, the “ashes” of the process, the fission products that are radioactive, constitute a, relatively, small volume. Yes, they have to be safeguarded and be kept away from the biosphere for thousands of years, but the volume of wastes is considered manageable.

Today (2011), there are 441 nuclear power plants operating in the world amounting to a total generating capacity of 375 GWe. The top five countries, in terms of number of operating plants are the USA (104), France (58), Japan (54), Russia (32), and Korea (21). In 2010, there were 62 plants under construction amounting to a capacity of 60.2 GWe. The top countries in numbers of plants under construction are China (25), Russia (11), Korea (5), and India (4); two plants under construction are in the USA, Japan, Bulgaria, Slovakia, and Ukraine.

- Today, $\sim 16\%$ of the electricity worldwide is produced by nuclear power plants; that fraction will

increase in the coming years as developing economies (e.g., China and India) and developed ones (Korea, Japan, and Europe) complete their announced ambitious nuclear expansion in order to satisfy their ever-increasing demand for electricity. The World Nuclear Association (WNA) projects that by 2060 at least 1,100 GWe of new nuclear capacity will be added.

What drives this “Nuclear Renaissance”? There are several factors:

1. *Increasing energy demand*: Due to an ever-increasing population and desire of the underdeveloped countries to improve their standard of living. In addition, a need for new plants materializes because old plants reach the end of their life and must be shut down (all types of electricity generating plants, not only nuclear, are designed with a finite operational life).
2. *Climate change concerns*: Increased awareness that fossil fuels release a large amount of greenhouse gases that may lead to a planetary climate change are driving decisions for new plants to be “green,” that is, to emit reduced amounts of greenhouse gases and other pollutants or not at all. Nuclear power plants are the only ones generating electricity with, essentially, zero emissions.
3. *Economics*: The main cost component of a nuclear plant is its construction cost; once this cost is overcome, the other two cost components for the generation of electricity (O&M and fuel) favor nuclear over fossil plants. Experience during the past 50 years, especially in the USA, has shown that nuclear is the best plant for generation of base load electricity, both in terms of cost and reliability.
4. *Fuel price stability*: Because the fuel cost constitutes a, relatively, small fraction of the total cost of generating electricity (for nuclear the main cost is the construction cost; fuel cost is between 10% and 16% of the total cost), even if the price of fuel doubles, the effect on the cost of electricity will be minimal. It should be noted that the cost components of the fuel are uranium, conversion, enrichment, fabrication-transportation; of these, only the uranium price fluctuated during the previous

50 years; the cost of the other components has been remarkably stable.

5. *Security of Fuel Supply:* Supply of fossil fuels (oil, gas, and coal) is vulnerable to interruptions of supply due to political turmoil, strikes, weather, etc. On the contrary, uranium is plentiful and available at reasonable prices, for the foreseeable future; more than that, since nuclear plants refuel every 2 years, it is quite likely that short-term upheavals can be settled before they have an effect on nuclear fuel supply.
6. *Nuclear Safety and Public acceptance:* The safety record of nuclear power plants during more than 50 years of operation (1957-2011) is outstanding. There were three accidents during that period. The TMI accident in the USA in 1979: it was a financial loss to the company operating the plant; not a single injury resulted to any person in or out of the plant and there was no contamination of the environment. The Chernobyl accident in 1986 was a very serious accident: radioactivity was released to the environment and 33 persons are known to have died, mostly firefighters. Some cancers will develop as a result of exposure to the radiation from the accident, making it a difficult task to quantify them. One should note, however, that no plant licensed anywhere today can have the Chernobyl type of accident; it is physically impossible (the Chernobyl design has been abandoned). Considerable area around the plant has been contaminated, but the reports today (2011) indicate that life is returning to normal. On March 11, 2011 a 9.1, Richter scale, earthquake hit Japan and affected the Fukushima nuclear site. There are six reactors on that site. On that day, two reactors were shut down and were not affected by what followed. When the quake hit, the four operating plants shut down, as design dictated. Shutting down means that the fission reaction is stopped; however, heat continues to be generated in the core from fission products, hence cooling must be provided. As per design, cooling continued using emergency diesel generators required by regulation and readily available for such eventuality. Then, 15–20 min later, the tsunami arrived; the waves of the tsunami swept away generators and their fuel supply and cooling of the core was lost. Fuel meltdown

of the cores occurred and radioactivity was released to the environment. Two plant workers were killed by the tsunami; these are the only immediate deaths as a result of this event. The tsunami itself and the earthquake resulted in the death of ~22,000 people. The area around the site was contaminated. Definitely the accident was a tremendous financial loss to the company operating the Fukushima plants.

As a result of the accident at Fukushima, two plant workers were killed by the tsunami; these are the only immediate deaths as a result of this event. At Chernobyl, ~33 persons, mostly firefighters, were the reported immediate deaths. Of course as a result of TMI, no deaths occurred; in fact in the USA there are zero deaths from the operation of commercial nuclear power plants from 1957 until today. Compare this, for example, to coal mining fatalities which amount to ~30/year in the USA; in other countries, e.g., China, the direct annual death toll from coal mining is much higher.

Without trying to deemphasize the importance of these three accidents, the net effect is that the world community still recognizes and accepts the generation of electricity using nuclear fission reactors as an indispensable component in the energy portfolio. At about 9 months after the Fukushima accident, there is no report of any country changing nuclear energy policy because of that accident (with the exception of Germany that decided to accelerate the shutdown of its nuclear fleet following a decision taken earlier by its Government). Overall acceptance of nuclear technology is the result of its excellent performance; another factor in favor of nuclear power is the lack of greenhouse gases by this technology or of any environmental effect in the vicinity of the plant. In the USA, acceptance of nuclear power by communities around the plant is ~70% or more favorable.

The nuclear industry, especially in the USA, learned and improved a lot after the TMI accident. I have no doubt that the international nuclear community will apply the lessons learned from Fukushima to make the already safe record of this industry even safer.

This section on Nuclear Energy consists of 19 articles that cover all aspects of the nuclear enterprise. Here is a brief description of each article:

- A. *Radiation sources*: It discusses radioactivity and basic radiation sources.
- B. *Radiation detection devices*: One characteristic of radioactivity is the fact that it can be detected relatively easily and accurately. This article discusses the devices used to accomplish such measurements.
- C. *Dosimetry and Health Physics*: Very early in the twentieth century (1920s), it was realized that ionizing radiation may be harmful to humans, therefore measures must be taken to protect people. These measures were based on (a) quantifying the effects of radiation exposure by establishing units of radiation dose and means to measure it and (b) establishing professional bodies that set protection standards [ICRP (1928), NCRP (1964) etc.]. The field of Health Physics was thus born resulting in great benefit for the radiation workers.
- D. *Fission reactor physics*: Fission reactors are the major sources of production of radioactive materials. What are the principles of their operation? What are the foundations of their safe operation? These are two of the major items discussed in this article.
- E. *Nuclear fuel cycles*: Providing nuclear fuel for a fission reactor is not a simple or straightforward task; it involves many steps (U procurement, conversion, enrichment, fuel rod and assembly fabrication). The users of the fuel are presented with choices, such as discarding the irradiated fuel as waste or reprocessing and recycling it. Also, reactor designers may affect the nuclear fuel “cycle” by building reactors that just produce electricity, or combine electricity production with generation of new fuels (breeders), or generate electricity in combination with burning some of the nasty by-products of the fission process. These are the matters discussed in this article.
- F. *Uranium reserves and mining*: How much uranium is there on our planet and at what price? Where is it found? How is it extracted? These are the topics of this article.
- G. *Nuclear fission power plants*: Once the fission reactor is designed and ready to operate, how is the fission energy utilized to generate electricity? The reactor core itself is not enough; plenty of other components must operate for the successful transformation of energy released in fission to electricity feeding a lightbulb. The fission reactor core makes a small part of a nuclear power plant. It is this aspect of nuclear power, components and activities outside the core, that is described in this article.
- H. *Nuclear reactor materials and fuel*: For a successful and long-term safe operation of a nuclear power plant, the materials used, especially those directly tied to the fuel, must function as designed (as expected) in the very hostile environment of a nuclear fission core. This article describes the pros and cons of the various materials that have been considered and the final choices made.
- I. *Nuclear safeguards and proliferation of nuclear weapons materials*: Of great concern to human kind is the acquisition of nuclear materials by groups or governments that may use them to make nuclear weapons, contrary to international treaties. This “proliferation” or rather “nonproliferation” of nuclear materials and possibly weapons is a concern that will never disappear; all that can be done by the international community of nations is to set up treaties, policies, and procedures that diminish the probability of proliferation. It is these aspects of this terrible problem facing humanity that are discussed in this article.
- J. *Radiation shielding and protection*: This article is complementary to articles A–C. Having discussed radiation sources, dangers from radiation, and standards of protection, how does one provide the means for a safe radiation environment for the workers and the public? How are relevant computations performed? Measurements? How is an effective radiation shield designed? These are the questions answered in this article.
- K. *Isotope separation methods for nuclear fuel*: The two elements found in nature that may be considered as fuels for fission reactors are Th and U. Unfortunately, only certain isotopes of these elements or made with the help of these elements can be manufactured into fuels. Hence, isotope separation methods must be employed for the

production/concentration of the “useful isotope(s).” These methods are discussed in this article.

- L. *Reprocessing of nuclear fuel*: Irradiated fuel contains many useful isotopes, primarily U and Pu. Reprocessing is the operation that extracts the useful isotopes from the irradiated (spent) fuel. The reprocessing methods used until today and those under research and development are discussed in this article.
- M. *Decommissioning of nuclear facilities*: Every nuclear facility has a finite lifetime; at the end, when operations stop for good, the law says that the site must, eventually, be returned to its preoperational state in terms of the presence of radioactivity. This is what decommissioning means. All the tasks associated with decommissioning are discussed in this article.
- N. *Radioactive waste management*: Storage, transport, disposal. The operation of fission reactors results in the production of radioactive materials. Such materials, if they have no further use, they must be safeguarded for long periods of time in order that their release to the biosphere may be prevented. The method of eventual disposal of such materials considered today by their producers is placement in a geologic repository. In the meantime, radioactive wastes must be stored and transported. These activities are discussed in this article.
- O. *GEN-IV reactors*: By any measure, current fission reactor designs are successful. However, there is room for improvement in terms of fuel utilization, thermal efficiency, use of the heat generated by the energy released in fission, multifunction of a nuclear plant, etc. There is considerable global effort underway to design fission reactors that will show some, if not all, of the improvements just mentioned. These new designs, collectively named as GEN-IV reactors, are described in this section.
- P. *Nuclear fusion*: Fusion reactors offer many advantages over fission reactors. Unfortunately, although fusion became known to man before fission and life on earth owes its existence to a fusion reactor in the sky (our Sun), no fusion plant has been built yet; fusion reactors present some unique challenges/difficulties that have not been resolved yet. But, the world’s scientific community is working as

a team in an effort to resolve the issues and build an operational fusion plant sometime in the future. All the past and current efforts in fusion research and the expected future developments are presented in this article.

- Q. *Nuclear power economics*: In a free market, every plant generating electricity must compete economically with all other options; nuclear is no exception, of course. Nuclear power presents some unique problems, with respect to financing, and these are the problems (and possible solutions) discussed in this article.
- R. *Thorium – An excellent ‘fertile’ nuclear fuel*: In addition to Uranium, Thorium (Th) is the only other element found on earth that can be used as a fuel in fission reactors. The Th properties, relevant to fission reactors, resources, availability and prices are discussed in this article.

Nuclear Facilities, Decommissioning of

DAVID R. TURNER

Geosciences and Engineering Division, Southwest
Research Institute, San Antonio, TX, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Types of Nuclear Facilities
General Methodologies for Decommissioning
Status of Decontamination and Decommissioning
Future Directions
Bibliography

Glossary

Decommissioning Nuclear decommissioning is a term used to describe the process of removing a nuclear facility or site safely from service and reducing residual radioactivity to a level that permits (1) release of

the property for unrestricted use and termination of the license or (2) release of the property under restricted conditions and termination of the license. Although waste classification and management is an important aspect of decommissioning, the details of radioactive waste management and disposal are not addressed in this article.

Decontamination The removal of undesired residual radioactivity from facilities, soils, or equipment, prior to the release of a site or facility and termination of a license. Also known as remediation, remedial action, and cleanup.

Exposure pathway The route by which radioactivity travels through the environment to eventually cause radiation exposure to a person or group.

Financial assurance A guarantee or other financial arrangement that ensures funds for decommissioning will be available when needed.

Institutional controls Administrative and physical measures to control access to a site and minimize disturbances to engineered measures established to control the residual radioactivity.

Monitoring The measurement of radiation levels, concentrations, surface area concentrations, or quantities of radioactive material and the use of the results of these measurements to evaluate potential exposures and doses.

Nuclear fuel cycle Consists of the different stages necessary to produce nuclear power. Specific stages include (1) the *front end* of the nuclear fuel cycle where uranium is mined and fuel is prepared, (2) the *service period* in which the fuel is used during reactor operation, and (3) the *back end*, which involves safe management, containment, and either reprocessing or disposal of spent nuclear fuel. Because uranium fuel is the most common type of nuclear fuel, this article focuses on the uranium nuclear fuel cycle.

Radiological survey An evaluation of the radiological conditions and potential hazards at a site related to the production, use, transfer, release, disposal, or presence of radioactive material or other sources of radiation. Radiological surveys can be used to provide the basis for acquiring necessary technical information to develop, analyze, and select appropriate cleanup techniques.

Residual radioactivity Radioactivity in structures, materials, soils, groundwater, and other media at a site resulting from activities under the licensee's control, excluding background radiation.

Definition of the Subject

The process of safely shutting down, dismantling, cleaning up, and monitoring nuclear facilities is collectively known as nuclear decommissioning. Nuclear power has been used as a source of energy for more than 50 years, and more than 500 nuclear reactors have been constructed and operated worldwide [1]. In addition to power plants, the nuclear fuel cycle requires different types of facilities to mine uranium, produce fresh nuclear fuel, and manage spent nuclear fuel and associated radioactive wastes after the fuel can no longer be effectively used to produce power. Many of the facilities associated with the nuclear fuel cycle that supports these reactors, as well as the reactors themselves, have either reached or are approaching the end of their planned service life. Also, some countries such as Belgium and Germany have initiated national policies to phase out nuclear power over time [2, 3]. Owners and operators of nuclear facilities, as well as government agencies responsible for their regulation, must evaluate economic and public policy considerations to determine whether to renew facility licenses or to permanently remove facilities from service, and decommission them to release the sites for other potential uses.

Definitions of nuclear decommissioning, radiological dose limits for site release, and even terminology can vary from country to country depending on the nature of the laws and regulations that govern the nuclear fuel cycle. The International Atomic Energy Agency (IAEA) defines decommissioning as "...the administrative and technical actions taken to allow the removal of some or all of the regulatory controls from a facility" [4]. Similarly, the Nuclear Energy Agency (NEA) broadly defines decommissioning as covering "all of the administrative and technical actions associated with cessation of operation and withdrawal from service" [1]. The World Nuclear Association (WNA) defines the two main objectives of decommissioning as rendering the site permanently safe and restoring it, "as far as practicable," for reuse for nuclear or non-nuclear activities. Reuse can apply to different components of a nuclear

facility, including land, water bodies, buildings, equipment, and materials [5].

Some early nuclear facilities were developed without explicit consideration of decommissioning, and these “legacy” sites continue to be identified and cleaned up – often at public expense. As it is currently practiced in most nations with a robust legal and regulatory framework, the decommissioning process begins when a facility is removed from service. Planning for decommissioning should begin at the design stage, before a facility enters operation [4, 6]. For example, early decisions about construction materials, site layout, spill prevention and control measures, waste management, and financial assurance can all influence decommissioning activities at the end of a facility’s life cycle and may be made (or be required) before a facility receives an operating license. Similarly, some national nuclear regulatory programs require that decommissioning plans be submitted with the initial license application and periodically reevaluated and updated during the operating life of the facility [6, 7]. Decommissioning is generally considered complete when the facility is removed from regulatory control (e.g., the license is terminated) and the site is made available for reuse.

The processes and technologies are similar to those used for other industrial facilities, but because of the nature of the materials used in the nuclear fuel cycle, nuclear decommissioning tends to be a regulated process that requires special procedures to handle and dispose of radioactive materials safely. The specific methods used in nuclear decommissioning vary widely, depending on factors such as the type, size, age, operational history, and design of a given nuclear facility. National policies differ on detailed objectives, and individual countries are likely to have different issues of concern such as the future use of nuclear power, the continued availability of trained staff, socioeconomic effects on surrounding communities that may result from shutting down a large facility, and financial issues associated with funding decommissioning activities.

For these reasons, there is no unique or preferred “one size fits all” approach to the nuclear decommissioning process. The intent of this article is neither to endorse any particular approach, nor to describe all potential aspects of nuclear decommissioning in detail, but rather to provide a

broad overview of generally applicable principles. Further, this article focuses on decommissioning facilities associated with the commercial-scale production of nuclear power. Other processes that use or generate nuclear materials such as military programs, nuclear medicine, and industrial operations that produce radioactive materials as a byproduct are beyond the intended scope of this discussion. The reader is referred to the references identified in the Bibliography for further information and more detailed discussion of the processes described here.

Introduction

In 2007, nuclear reactors provided slightly more than 14% of the world’s electricity, with ranges as high as 76.8% in France [8]. Currently (July 2009), 436 nuclear power reactors are in operation, with 5 reactors in long-term shutdown and 48 new reactors under construction [9]. In addition to nuclear reactors, generating nuclear power requires different types of facilities to support the nuclear fuel cycle. For example, as of August 2003, the IAEA reported that there were 423 operating nuclear fuel cycle facilities, with 19 more under construction. Eventually, decisions will need to be made with regard to closing and decommissioning these facilities in a safe manner and potentially returning the land to other uses.

More than 50 years after the first nuclear power reactor went on line, many of the facilities associated with commercial-scale nuclear power generation are now approaching the end of their planned service life. For example, the European Union estimates that at least one-third of the 152 nuclear power plants operating in its member countries will need to be decommissioned by 2025 [2], and the IAEA reported that 297 nuclear fuel cycle facilities were shut down or in the process of being decommissioned worldwide as of August 2003 [4]. An even larger number of manufacturing and research facilities (about 1,600) that use radioactive material will need to be decommissioned “over the coming decades” [10].

At the same time, increased global energy demand, coupled with a growing concern about the effects of carbon emissions from traditional fuel sources, has sparked renewed interest in nuclear power generation. As an example, the U.S. Nuclear Regulatory

Commission (USNRC), the agency responsible for licensing and regulating commercial nuclear activities in the United States, received a total of 17 applications to construct and operate 26 new commercial nuclear reactors during 2007–2008 [11]. Although some countries are deemphasizing nuclear power, other countries in Europe and Asia have indicated a renewed interest in nuclear power as a component of their overall energy portfolio [3, 8, 12].

Together, these developments indicate that all stages of the commercial nuclear fuel cycle will continue to be active or will be expanded in the foreseeable future. For this reason, methods to safely take these facilities out of service and decommission them will continue to be an important component of energy policy. It is these methods that are the subject of this article.

Although primarily developed for nuclear power plants, there are three basic types of alternative decommissioning strategies that may be applicable to other nuclear facilities:

- A strategy of immediate decontamination and dismantlement [1, 7] (also defined as DECON [11, 13]) begins soon after the nuclear facility closes.
- Safe storage [1], deferred or delayed decontamination [7], or SAFSTOR [14, 15] refer to decommissioning strategies where a nuclear facility is left intact after closing, placed in a stable condition, and maintained and monitored until subsequent dismantlement and decommissioning. Similarly, uranium production facilities (mines and mills) may be placed on standby status when uranium prices are low and deposits cannot be produced at a profit.
- A strategy of entombment [1, 7] or ENTOMB [14, 15] involves encasing radioactive materials onsite in a long-lived, structurally sound material such as concrete.

The first two strategies may also be combined. For example, some facilities at a site may be immediately dismantled while other structures are placed in safe storage. Generally, decommissioning activities are anticipated to be completed in a period of decades from the end of operations [14, 15].

Once the decommissioning strategy is selected for a given facility, general activities associated with nuclear decommissioning may include:

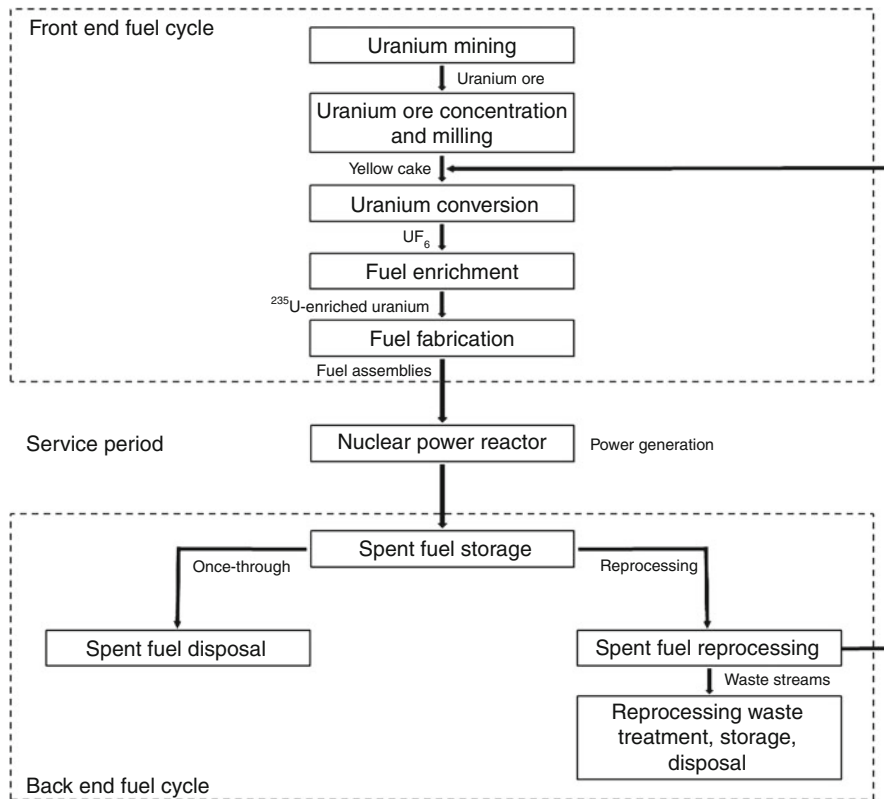
- Characterizing the features of the site and conducting radiological surveys to determine radiation background and residual radiation levels
- Developing a site-specific decommissioning plan
- Estimating cost
- Conducting safety and performance assessments
- Decontaminating structures, equipment, and components for reuse or recycling
- Dismantling and removing buildings, structures, and equipment from the site
- Remediating contaminated soils and groundwater
- Performing waste management and disposal
- Conducting final inspections and surveys
- Reclaiming disturbed lands
- Implementing active institutional controls and monitoring

Specific decommissioning activities and technologies are determined on a case-by-case basis and can depend on many things at a given site, such as the duration, type, and scale of operations; the geologic setting of the site; socioeconomic considerations; and the regulatory policies of the government. In addition, a key part of a decommissioning plan is estimating decommissioning costs to establish financial arrangements that ensure resources are available to complete the decommissioning process.

Types of Nuclear Facilities

As described in the "Introduction" section, generating electricity from nuclear power plants is only part of the nuclear fuel cycle [16] and decisions will need to be made with regard to the safe closure and decommissioning of each of these facilities. To understand any unique aspects that will need to be addressed during nuclear decommissioning, it is important to describe the general nature of the activities that are conducted at each type of facility in the nuclear fuel cycle. Because uranium fuel is the most common type of nuclear fuel, this article focuses on the uranium nuclear fuel cycle (Fig. 1).

To support the nuclear fuel requirements of commercial power-generating facilities, there are facilities that



Nuclear Facilities, Decommissioning of. Figure 1

Simplified diagram representing the nuclear (uranium) fuel cycle

- Produce uranium by mining
- Mill ore to concentrate the uranium and package it for transportation
- Purify and transform the uranium concentrate into a form suitable for fuel manufacture through a process called uranium conversion
- Enrich the uranium in isotopes (^{235}U) that produce sustainable nuclear reactions
- Fabricate reactor fuel components and fuel assemblies

These facilities are part of the “front end” of the nuclear fuel cycle. Decommissioning issues for the front end of the fuel cycle are typically associated with naturally occurring radioactivity, such as uranium and radium, and hazards associated with the chemical processing of natural uranium-bearing ores [5].

After the nuclear fuel is used in a commercial reactor to produce electrical power during the service period, the management of the spent nuclear fuel at

the “back end” of the nuclear fuel cycle can include facilities that

- Reprocess the fuel to extract nuclear materials and recycle uranium back into the front end of the fuel cycle
- Store or permanently dispose of spent nuclear fuel

At the back end of the fuel cycle, high-level sources of radioactivity and direct irradiation from spent nuclear fuel are decommissioning concerns in addition to natural radioactivity and chemical processing hazards [5]. An example of the typical material balance for the annual operation of a 1,000 Megawatt electric (MWe) nuclear power reactor [17] is included in Table 1.

Each of these facilities will require nuclear decommissioning at the end of its life cycle. Usually, the operator of a facility develops a decommissioning plan (see “[Developing a Site-Specific Decommissioning Plan](#)”) to identify specific activities, estimate costs, and

Nuclear Facilities, Decommissioning of. Table 1 Material balance for the annual operation of a 1,000 MWe nuclear power reactor [17]

Type of nuclear facility	Material balance ^a
Mining	20,000 metric tons (22,000 t) of 1% uranium ore
Milling	230 metric tons (250 t) uranium oxide concentrate (U ₃ O ₈) containing 195 metric tons (215 t) of uranium
Uranium conversion	288 metric tons (317 t) uranium hexafluoride (UF ₆)
Fuel enrichment	35 metric tons (39 t) UF ₆ , containing 24 metric tons (27 t) enriched uranium (4% U-235), 11 metric tons (12 t) depleted uranium (0.25% U-235) tails
Fuel fabrication	27 metric tons (30 t) UO ₂ , containing 24 metric tons (30 t) enriched uranium
Reactor operation	8,640 million kilowatt-hour electricity at full output (assuming 100% load factor)
Spent nuclear fuel	27 metric tons (30 t) spent nuclear fuel containing 23 metric tons (25 t) uranium (0.8% U-235 as UO ₂), 240 kg (529 lb) plutonium, 720 kg (1,587 lb) fission products, and transuranic elements

^aAssuming enrichment to 4% U-235 with 0.25% tails assay; core load 72 metric tons (79 t) U, refueling so that 24 metric tons (26 t) U/year replaced; operation – 45,000 MWday/t (45 GWday/t) burn-up, 33% thermal efficiency

lay out a schedule [18–21]. The plan is submitted to the regulatory agency to ensure that it complies with the applicable regulations. Once it is approved, the initial decommissioning plan is periodically updated and becomes more detailed as the facility evolves during its operational life. At the end of the facility life cycle, the decommissioning plan serves as a blueprint for the nuclear decommissioning process. During decommissioning, the regulatory agency may periodically inspect the site to ensure that the decommissioning plan is being implemented correctly.

The following sections introduce the types of nuclear facilities associated with generating nuclear power and identify features that may influence nuclear decommissioning decisions and activities. The listing is not intended to be exhaustive, nor is it intended to provide a detailed discussion of the complex regulatory controls that may apply for a given type of nuclear facility. The reader is referred to the references identified in the Bibliography for further reading.

Front End Fuel Cycle Facilities

Uranium Mining At the start of the nuclear fuel cycle, uranium mining focuses on extracting natural uranium ore from the earth. Uranium mining may be done through a conventional process of excavation by

open pit or underground mining techniques. During excavation, uranium ore is segregated from waste rock or overburden, and shipped to a uranium mill for further processing (see “Uranium Milling”). Open pit and underground uranium mines are typically decommissioned and reclaimed in accordance with regulations applicable to the mining industry in general [22]. In many ways, decommissioning a uranium mine is subject to challenges similar to those faced in cleaning up mining operations for other resources such as coal and metals. For example, excavations may need to be backfilled, pit walls and disturbed surfaces recontoured, and revegetated to meet applicable mine reclamation standards. Groundwater contamination plumes with elevated levels of uranium and associated heavy metals (e.g., arsenic, selenium) need to be remediated and monitoring systems installed, as appropriate.

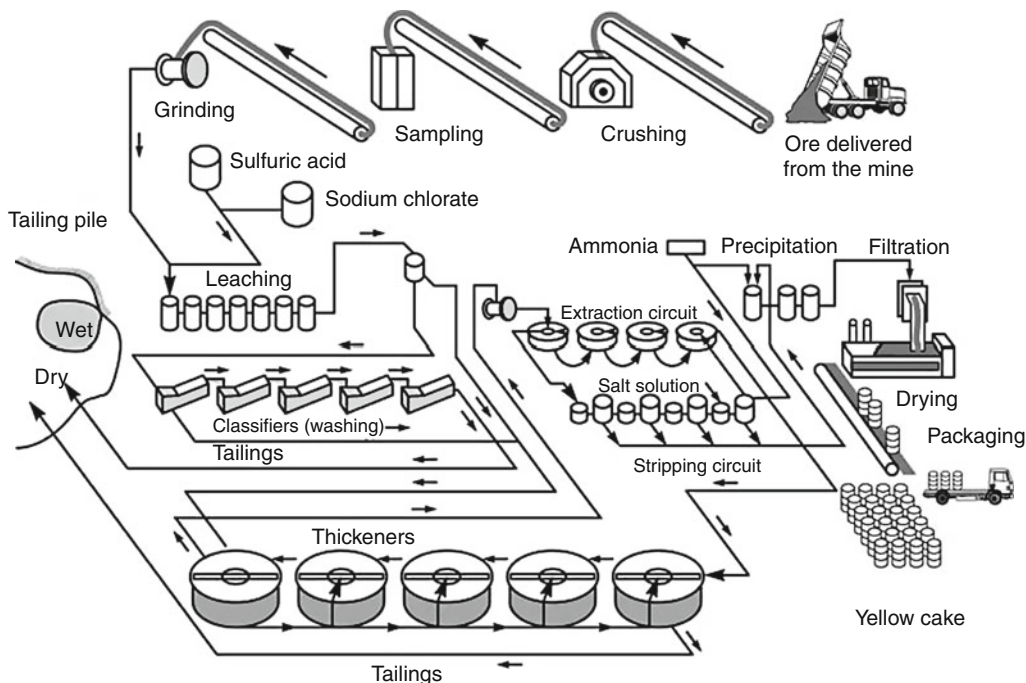
Alternatively, uranium recovery may be through a process called in situ leaching (ISL), where chemical fluids are injected through a series of wells into the subsurface to dissolve uranium from ore minerals. The now uranium-enriched solution is pumped back to the surface for subsequent extraction and processing [13, 22–24]. Decommissioning of ISL uranium facilities is different from cleaning up conventional uranium mining operations. For example, because there are no

large-scale excavations associated with the ISL uranium recovery technology, surface land disturbance is much less than with conventional mining methods, and large amounts of waste rock are not generated. This can simplify the reclamation effort, but impacts to groundwater are potentially greater, and restoration of groundwater quality to premining levels tends to be a focus in decommissioning ISL facilities [13]. In addition, although the surface facilities such as well heads and pump houses necessary to support ISL uranium mining may be small compared to conventional operations, the well fields themselves may be very large. For example, the permitted areas for U.S. ISL operations in Wyoming and Texas may be as large as 6,500 ha (16,000 acres) and individual well fields may contain hundreds to thousands of wells [13].

Uranium Milling While varying depending on the deposit and the type of ore, uranium ores typically have a concentration of about 1% or less uranium (U_3O_8) by weight, although it can be as high as 20%. Decommissioning uranium milling facilities is generally concerned with naturally occurring radioactive

elements such as uranium, thorium, and radium. For conventional uranium mining and milling operations (Fig. 2), uranium is concentrated through milling (crushing, grinding, separation) and subsequent chemical processing of the ore using alkaline or acid leaching solutions to produce a product called “yellow cake,” a coarse powder that is approximately 70% or more U_3O_8 by weight. Milling operations are similar for ISL uranium, but because the mill input is uranium-bearing solutions instead of solid ore, the crushing and grinding, and leaching circuits are not necessary.

Depending on the economics of the operation, a mill may be colocated with a uranium mine, or it may be separate, charging a “toll” to accept and mill ore (conventional) or uranium-bearing solutions (ISL) from nearby uranium mines. Because of the throughput of uranium-bearing solutions and the handling of yellow cake, some of the buildings, pipes, and equipment in the mill may become radioactively contaminated with natural radioactive elements over time and will need to be decontaminated and/or disposed of during decommissioning. Spills and unintentional releases of uranium-rich solutions can lead to soil and



Nuclear Facilities, Decommissioning of. Figure 2

A typical conventional uranium mill [22]

groundwater contamination that may need to be remediated during decommissioning. In addition, during conventional milling, large quantities of waste solids are produced by the crushing, grinding, and leaching circuits. These waste solids, or “tailings,” are placed in large impoundments that will also need to be decommissioned at the end of the mill’s operational life. ISL facilities do not treat solid rock and therefore produce no tailings. Because these facilities handle large volumes of uranium-rich solutions, however, they may have one or more evaporation ponds that will eventually need to be decommissioned. Although specific activities will vary depending on size, age, operational conditions, and design, decommissioning of a uranium mill (both for conventional and ISL facilities) includes general activities described in the “General Methodologies for Decommissioning” section.

Uranium Conversion Although natural uranium ores are concentrated and purified during milling to produce yellow cake, the uranium must be further purified and converted to a gas state to produce nuclear fuel. The nature of these chemical processes and the size and complexity of these uranium conversion facilities strongly influence the decommissioning process. Uranium conversion facilities are designed to remove impurities from the yellow cake and convert solid uranium into uranium hexafluoride (UF_6), the only uranium compound that exists as a gas at suitable temperatures [4, 17]. The UF_6 gas is then pressurized and cooled to a liquid state [4]. The liquid is stored in cylinders and allowed to solidify for shipment to a fuel enrichment plant (see “Fuel Enrichment”). Some uranium conversion facilities also recycle uranium scrap that may come from manufacturing facilities or other production plants [4].

The uranium conversion process involves strong acids and alkali agents to dissolve the yellow cake powder. As with mining and milling, the operational risks associated with uranium conversion are chemical as well as radiological, and the safe removal and disposal of hazardous chemicals used in the conversion process must be taken into consideration as part of the decommissioning process. Because it is predominantly an industrial chemical facility, the types of decommissioning issues for a uranium conversion

facility are similar to those described in the “Uranium Milling” section for buildings, equipment, and land reclamation at uranium mills. No tailings are produced during uranium conversion. Groundwater and soil contamination from spills and leaks are possible and will need to be identified, remediated, and monitored in accordance with applicable regulations [19]. The size and complexity of uranium conversion plants can lead to significant decommissioning costs. For example, alternative decommissioning options for the 243-ha (600-acre) Sequoyah Fuels uranium conversion plant site in Gore, Oklahoma, varied from about US \$19 to US \$254 million (2006 dollar value), depending on the option selected for disposal of contaminated materials [25].

Fuel Enrichment Natural uranium is composed of ^{238}U (99.274%) and ^{235}U (0.711%) and contains trace amounts of ^{234}U (<0.01%). Nuclear reactor fuel requires increasing (or enriching) the fissionable ^{235}U above these natural levels to sustain the nuclear reactions. During the fuel enrichment process, gaseous UF_6 is gradually enriched in ^{235}U to about 3–5% – levels that are sufficient for fabricating commercial reactor fuel [26, 27]. One product of the enrichment process of special concern during decommissioning is depleted uranium tails, where the ^{235}U content has been reduced to below natural levels (about 0.2–0.35%) [28]. This waste stream is commonly stored at the site in gas cylinders of UF_6 , although it may also be converted to solid form such as uranium oxide. Depleted uranium emits low levels of radiation and is also a toxic heavy metal, so during decommissioning, this waste must be managed and disposed in accordance with both radiological and hazardous waste criteria [27, 29].

Gaseous diffusion is the process most widely used for nuclear fuel enrichment. The gaseous diffusion process takes advantage of the different diffusion rates that result from the slight differences in mass for ^{238}U and ^{235}U . Gaseous UF_6 is pumped through a permeable porous barrier media. The lower molecular weight ^{235}U has a higher diffusion rate and moves through the barrier media more readily than the higher molecular weight ^{238}U ; the UF_6 gas that passes through the media is therefore slightly enriched in ^{235}U [30]. The process

is repeated through many barriers until the desired enrichment levels are achieved.

These types of plants are large industrial facilities, with a footprint of about 300–600 ha (750–1,500 acres) that contains a large amount of piping and pumps required to move the UF_6 gas through the permeable barrier system [4]. During operations, the primary hazards in gaseous diffusion plants that may influence subsequent decommissioning include the chemical and radiological hazard of a UF_6 release [31]. There is also a potential for mishandling the enriched uranium, which could create a criticality accident (inadvertent nuclear chain reaction). Because these are large facilities, decontamination and decommissioning of inactive buildings and areas may occur while other parts of the facility continue to operate. Surveillance, maintenance, and security will continue for active parts of the facility. Depending on the operating history of the plant, including spills and unintentional releases, decommissioning and cleanup activities at fuel enrichment facilities may include assessing and remediating soil or groundwater and waste management activities, such as disposing of contaminated materials. The size and complexity of a gaseous diffusion plant can lead to large costs for full decommissioning of all facilities. For example, the U.S. Government Accountability Office (GAO) estimated that cleanup activities at three fuel enrichment plants (Paducah, Kentucky; Oak Ridge, Tennessee; Portsmouth, Ohio) cost US \$2.7 billion (in 2004 dollars) from 1993 through 2003, and total costs through final decommissioning in 2044 would exceed revenues into the Uranium Enrichment Decontamination and Decommissioning Fund by about US \$3.2–6.2 billion (in 2007 dollars) [32]. Newer technologies such as gas centrifuge and laser separation are being considered for the next generation of fuel enrichment facilities, and the decommissioning issues are likely to be different, with perhaps less waste generated during operation [26].

Fuel Fabrication For a typical commercial light water reactor, nuclear fuel is the solid form of uranium oxide (UO_2). Fuel fabrication facilities use chemical processes to convert the ^{235}U -enriched UF_6 into UO_2 in the form of a fine powder [28]. Because many of the materials (such as UF_6) are the same, specific factors that may influence the decommissioning of fuel

fabrication facilities will be similar to those identified for uranium conversion and fuel enrichment facilities (see “[Uranium Conversion](#)” and “[Fuel Enrichment](#)”). This powder is then compacted and sintered (heated at a high temperature to fuse the particles together) to produce fuel pellets. These pellets are loaded into metal tubes to produce fuel rods. Hardware is then used to configure the fuel rods into fuel assemblies of the appropriate dimensions and design for a nuclear power reactor. Although this article focuses on the uranium nuclear fuel cycle, other nuclear fuels can also be fabricated, including mixed oxide (MOX) fuels formed from combining uranium and plutonium oxides, thorium fuels based on the ^{232}Th decay chain with ^{233}U as the fissile fuel element, uranium metal alloy fuels, and microsphere fuel particles [28].

Heavy water (water that contains more than the natural proportion of the hydrogen isotope deuterium, ^2H) is used as a moderator in some types of nuclear reactors. Heavy water is extracted from normal water through several chemical processes, the most common of which is distillation through electrolysis or isotopic exchange [24].

Chemical, radiological, and criticality hazards at fuel fabrication facilities are similar to hazards at enrichment plants. Most at risk from these hazards are the plant workers. These facilities generally pose a low risk to the public.

Service Period: Nuclear Power Plants

As indicated, the three decommissioning strategies described in the “[Introduction](#)” section were initially developed for nuclear reactors and power plants. The principal concerns during the decommissioning of these reactors and power plants are the safe cessation of operation; the safe management, storage, and disposition of highly irradiated spent nuclear fuel; draining and treatment of water and other fluids from the reactor cooling systems; and the decontamination and disposal of equipment, materials, and other systems that may contain contamination or activation products at the site.

At a very basic level, most nuclear reactors operating today use the heat from the controlled fission of ^{235}U (and perhaps ^{239}Pu in the case of MOX fuel) to boil

water that turns a turbine and produces electrical power. As described previously, there are 436 nuclear power reactors operating worldwide, with 48 new reactors under construction. In addition, there are 287 (as of August 2003) research and test reactors and critical assemblies (i.e., producing little or no power), predominantly located at research universities and government facilities, that are used for research, education, and training purposes [4, 33].

There is no single design that is representative of all reactors. Of the operating nuclear power reactors, about 400 are water cooled and moderated (Energy Information Administration, 2006) and are predominantly pressurized water reactors (PWR), boiling water reactors (BWR), and pressurized heavy water reactors (PHWR). For example, all of the 104 operating commercial nuclear reactors in the United States are either BWR (35 reactors) or PWR (69 reactors) types [34]. Other operating reactor types include gas-cooled reactors, graphite-moderated reactors, and fast breeder reactors.

Early prototype nuclear reactors are sometimes called Generation I reactors. The current nuclear reactor designs are sometimes called Generation II (large central nuclear power plants) and Generation III (advanced LWR) reactors [28, 35]. Current research and development efforts are focused on designing the next generation, or Generation IV reactors, with a goal to provide more efficient and safe nuclear power generation that is also more resistant to nuclear proliferation [28, 36–38]. In the United States, the Next Generation Nuclear Program initiated with the Energy Policy Act of 2005 focuses on developing a very-high-temperature gas-cooled reactor (VHTR) operating at temperatures greater than 950°C for the production of electricity, process heat, and hydrogen [37]. Other designs are also being considered for the next generation of nuclear reactors, predominantly those based on gas-cooled (such as pebble bed modular reactors), water-cooled (super-critical water-cooled reactors), and fast-spectrum technologies (cooled by sodium, lead, or inert gases) [28, 35].

The different current and future design types and sizes of reactors make decommissioning inherently a site- and reactor-specific process. Each current and future reactor design will have its own design-specific decommissioning requirements that must be taken into consideration.

Back End Fuel Cycle Facilities

After the nuclear fuel is irradiated in a reactor and the useful energy has been extracted, it is called spent nuclear fuel. The removal of spent nuclear fuel from the reactor is generally considered part of the transition from the operational phase of the power plant, and not as part of the decommissioning process. The back end of the nuclear fuel cycle consists of facilities that handle spent nuclear fuel. Currently, most commercial nuclear power is based on an open, or “once through” uranium fuel cycle, where the fuel is used in a power plant one time and then removed as spent nuclear fuel [39].

Fuel Reprocessing Facilities When the spent nuclear fuel is removed from the reactor, it is predominantly composed of uranium oxide (96%), other actinides like plutonium and americium (1%), and other fission products such as cesium and strontium (3%) [30]. The spent nuclear fuel must be cooled both thermally and radioactively in a water-filled spent fuel pool and later placed in dry cask storage at the reactor site.

Fuel reprocessing facilities are designed to recover materials such as uranium and plutonium from irradiated spent nuclear fuel. After sufficient cooling, the spent nuclear fuel is dissolved using solvents and the usable components (mostly uranium and plutonium) are separated from waste materials such as other actinides and fission products [4, 40]. The recovered uranium and plutonium are recycled into the front end of the nuclear fuel cycle and refabricated to produce new nuclear fuel (such as MOX) or used for defense purposes. Waste materials, in the form of sludges, salt cake, or calcined wastes, and the reprocessing solutions are collected for disposal. Liquid wastes are generally not suitable for disposal, and decommissioning and cleanup may include vitrification to solidify waste solutions in the form of borosilicate glass.

Because of the high radiological dose rates and contamination levels associated with irradiated spent nuclear fuel, human access is limited for major parts of the facility. This leads to decontamination and decommissioning that is more complicated than facilities such as uranium mills and fuel fabrication facilities at the front end of the nuclear fuel cycle. Facilities tend to be very large, and large volumes of liquid wastes are produced and stored for subsequent disposal.

Examples of fuel reprocessing facilities in the United States include U.S. Department of Energy (USDOE) facilities at Hanford, Washington, and the Savannah River site in South Carolina. The only commercial fuel reprocessing facility in the United States, located at West Valley, New York, ceased operations in 1972. The Carter administration elected to defer reprocessing of commercial nuclear fuel in 1977, and there are no commercial fuel reprocessing facilities currently operating in the United States. The USDOE has completed vitrification of the liquid wastes at West Valley, storing the borosilicate glass logs on site [41, 42]. Decommissioning activities at West Valley are ongoing. Large reprocessing facilities are also located at Sellafield in the United Kingdom and La Hague in France.

Waste Management and Disposal As with many aspects of nuclear decommissioning, the nature and amount of waste produced during cleanup activities will depend on the age and size of a given facility, as well as the nature and history of operations at the site. As described in IAEA [43], four different types of waste or “waste streams,” each with different disposal options, are typically produced during nuclear decommissioning. The first three are types that include radioactive waste:

- Primary waste is generated during dismantling activities and may include internal plant components such as a reactor pressure vessel and associated piping, equipment such as pumps and valves, special facilities such as glove boxes and radiation hot cells, and building materials. These components of the primary waste streams may be radioactive through activation by both short- and long-lived radionuclides during plant operations, or by surface contamination. Primary wastes can include a variety of materials, but typically metal and concrete rubble are the largest component by volume of this waste stream [43].
- Secondary wastes are generated during different steps in decontamination and dismantling. These may include solutions, absorbents, and filters used to treat surfaces to reduce radioactive contamination.
- Tools and equipment, such as cutting equipment and protective gear for personnel use during

decontamination and dismantling, may become contaminated. To minimize costs and waste volumes, some equipment may be decontaminated so that it can continue to be used.

In addition to these radioactive wastes, decommissioning can potentially produce large volumes of non-radioactive wastes, such as construction debris, sanitary wastes, hazardous chemicals, and asbestos. These wastes are similar to those that might be encountered during the decommissioning of a typical, nonnuclear industrial facility. Some of these materials like furniture, nonirradiated scrap metal, and office equipment can be reused and recycled. The remainder of this waste stream typically has a well-established disposal path through municipal landfills and sanitary disposal systems.

Radioactive waste can occur in a variety of forms, including gases, liquids, and solids, with radiological characteristics that depend on the concentration and half-lives of the different radionuclides. For these reasons, classification of waste is particularly important, as it will determine the methods to be used in handling, segregating, conditioning, packaging, and transporting wastes. Waste classification also establishes the applicable acceptance criteria for different waste storage and disposal options, which in turn can have a strong effect on decommissioning planning and implementation, and ultimately, the total cost of decommissioning [43].

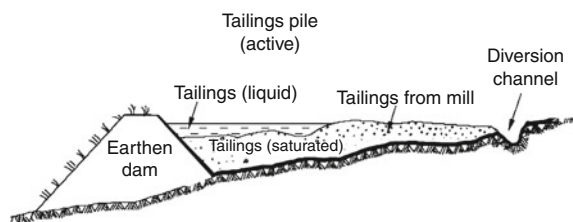
Although it varies from country to country, classification of radioactive waste is typically based on some combination of the types of processes from which the waste was generated, the radionuclide content of the waste, the timing of the waste generation (e.g., legacy wastes generated prior to developing a robust licensing process), as well as chemical, physical, and biological properties of the waste [30, 43, 44]. The following sections provide general descriptions based on waste classifications proposed in IAEA [45]. Note, however, that specific classification of waste streams and the methods for their management, treatment, and disposal are defined by the policies, laws, and regulations that a nation applies to commercial-scale nuclear power generation and subsequent nuclear decommissioning and waste management activities. For these reasons, waste classification is complex and may vary from country to country [44]; the descriptions presented

here are for informational purposes only and are not an endorsement of a particular radioactive waste classification system.

Uranium Mill Tailings and Groundwater Restoration For conventional mining and milling operations, processing the ore results in a waste stream of solid waste material, sands, fine-grained slurries (sometimes called slimes), and processing liquids. These materials are collectively referred to as tailings and are pumped from the mill to a tailings impoundment for disposal (see Figs. 2 and 3).

Generally, there are a number of tailings impoundments and evaporation ponds at each mill site [22], and the total volume can be quite large. The amount and nature of the tailings, however, depend on the capacity of the mill and the length of time the facility is in operation. For example, the former Climax Uranium Company mill in Grand Junction, Colorado, produced 2 million metric tons (2.2 million tons) of tailings that were placed in a 46-ha (114-acre) tailings impoundment [46]. The nature of the material in each impoundment or pond varies, depending on which part of the mill process produced the tailings. In addition, with the approval of the regulatory agency, operators may use tailings impoundments for onsite disposal of wastes associated with decommissioning and dismantling buildings, structures, and equipment, as well as for soils, pond liners, and sludges that have been radiologically contaminated.

Tailings have become a major focus in regulating active uranium mining and milling operations, as well as cleaning up legacy tailings from older, inactive sites. For example, in 1978, the U.S. Congress enacted the Uranium Mill Tailings Radiation Control Act (UMTRCA) to provide government funds to clean up and stabilize tailings from inactive legacy mills [22].



Nuclear Facilities, Decommissioning of. Figure 3
Schematic cross section of a tailings impoundment [22]

In a 1995 summary study of UMTRCA sites, the USDOE noted that tailings reclamation was the costliest aspect of the decommissioning process for conventional uranium mills [22]. During reclamation planning, the operator typically uses computer models to evaluate geotechnical stability. Although site-specific tailings reclamation methods are established in a reclamation plan that is evaluated for compliance with the government regulations that govern the cleanup [22, 24, 47, 48], general steps include:

- Shaping and recontouring the tailings pile and installing drainage diversion systems to minimize erosion hazards from surface runoff.
- Allowing the tailings to settle and dehydrate.
- Establishing survey monuments so that settling can be monitored.
- Installing a low permeability engineered cover to minimize water infiltration into the pile and radon emissions from the tailings. Generally this cover consists of clays and/or geotextiles.
- Installing a final cover for erosion protection.
- Establishing a monitoring system to ensure that the design and construction of the reclaimed tailings work as expected. Long-term stewardship during a period of institutional controls may include active measures to maintain and repair the tailings covers.

For active operations, regulations typically call for lining tailings impoundments and installing monitoring systems to prevent migration of contaminants (e.g., radionuclides and associated heavy metals) from the tailings into underlying aquifers [24, 48]. For older and inactive operations, however, the tailings impoundments may not be lined. This can lead to contaminants leaching into the underlying groundwater system over time; these contaminants need to be remediated when the tailings are reclaimed. Depending on the importance of local groundwater resources to nearby communities and the site and design of the impoundment, the level of effort needed to clean up groundwater may be extensive and long term.

Because uranium milling at an ISL facility does not produce tailings, the amount of material at the surface to be decommissioned and ultimately reclaimed is significantly less. Because there are no large tailings impoundments associated with ISL facilities, contaminated decommissioning wastes (e.g., equipment,

building components, soils, evaporation pond sludges, and liners) must either be transported to a licensed disposal facility, or an onsite disposal cell will need to be built.

The ISL process does, however, alter groundwater chemistry in the production well fields. After uranium extraction from the ore deposit is no longer economically feasible, the groundwater quality is restored to pre-extraction conditions through a series of treatment steps [13, 49]. As with tailings reclamation, site-specific methods are determined with approval of the regulatory agency, but general steps for groundwater restoration include:

- Groundwater sweep, where groundwater is pumped from the production zone to draw in surrounding natural groundwater. The pumped groundwater may be treated to remove contaminants and reinjected.
- Reverse osmosis, where groundwater is pumped from the production zone and passed through a pressurized, semipermeable membrane (reverse osmosis) system to remove dissolved chemicals. The cleaner water is then reinjected, and the more concentrated brines are pumped to evaporation ponds or disposed of by injection in deep wells.
- Stabilization, where chemicals such as hydrogen sulfide may be injected into the production zone to establish chemical conditions that cause dissolved chemicals to precipitate out as minerals.
- Monitoring, where the groundwater quality is tested to ensure that conditions have stabilized and meet the restoration criteria.

Groundwater restoration occurs when the economics for producing a well field are no longer favorable. Because ISL facilities typically have more than one well field, it is common for groundwater restoration to begin in one well field while other well fields are still actively producing uranium [13, 49].

After groundwater restoration, the physical components of the well field are then restored. Piping, well casing, and pipeline materials are hauled to a licensed disposal facility. Surface facilities such as tanks and buildings are dismantled, removed, and disposed. Pumps are removed for reuse in other well fields, and the wells are filled with cement, plugged, and abandoned. Any soils contaminated by well field spills are

removed and disposed, and the disturbed land is graded, recontoured, and revegetated.

Spent Nuclear Fuel and High-Level Waste As described previously in “[Fuel Reprocessing Facilities](#)” section, after the economic energy has been extracted from nuclear fuel, the fuel is removed from the reactor. At this point, it is referred to as spent nuclear fuel. Spent nuclear fuel is highly radioactive because of the decay of fission and activation products that result from the nuclear reactions that occur when the fuel is inside the reactor core. Although it is no longer hot enough to generate electricity, it is well above ambient temperatures.

For these reasons, spent nuclear fuel must be carefully handled, stored, and shielded to provide both radiation protection to workers and the public, and to manage the thermal heat generated by the cooling fuel. In most cases, spent nuclear fuel is stored at or near the reactor, either in dedicated spent fuel pools that use water to provide both cooling and radiation protection or in air-cooled concrete and steel dry casks [13].

As described in “[Fuel Reprocessing Facilities](#)” section, after spent nuclear fuel has been cooled and radioactivity has decreased through decay of short-lived radionuclides, spent nuclear fuel may be reprocessed. The fuel reprocessing process typically produces a liquid waste stream that contains fission and activation products that remain after potentially valuable radioactive elements such as plutonium and uranium are removed. This liquid waste stream is typically stored in large tanks for subsequent treatment. Treatment may include processing the liquid wastes into different forms such as sludges, salt cake, or a calcined solid. These forms still produce radiation and must be shielded to provide protection to workers. Current practice in countries such as France, Japan, the United Kingdom, and the United States involves using high temperature furnaces to vitrify liquid wastes into a solid glass waste form.

In most countries in Europe, Asia, and North America, the ultimate disposal path for spent nuclear fuel and reprocessing high-level radioactive waste (HLW) is permanent geologic disposal in an underground repository [50]. At present, a large number of different geologic settings are under consideration. No country has licensed and constructed a geologic

repository, however, and spent nuclear fuel and reprocessing HLW are typically managed through onsite interim storage.

Intermediate-Level Waste Intermediate-level radioactive waste (ILW) is defined in IAEA [45] as "... waste which, because of its radionuclide content requires shielding but needs little or no provision for heat dissipation during its handling and transportation." This waste may be further classified into components consisting of short-lived radionuclides that will decay to low levels during a period on the order of hundreds of years in which institutional controls such as fencing or access restrictions can be considered to be effective in minimizing radiological dose [45]. Conversely, long-lived ILW is dominated by radionuclides that will not decay to sufficiently low levels. The ILW classification is not used in the United States.

In the United States, a transuranic (TRU) waste stream includes man-made alpha radiation-emitting radionuclides with an atomic number greater than that of uranium (i.e., 92) and a half-life longer than 20 years [51]. TRU is produced during reactor fuel assembly, weapons fabrication, and fuel chemical processing operations [15]. Specifically, TRU is that portion of the waste stream that is not classified as spent nuclear fuel, HLW, or low-level radioactive waste (LLW) [30, 51].

A wide variety of storage options exists for the storage and disposal of ILW and TRU. Storage may be through retrievable burial, underground bunkers, concrete caissons, aboveground concrete pads, and inside buildings [30]. Since 1999 in the United States, the USDOE has been disposing TRU waste in a bedded salt deposit about 700 m (2,300 ft) below the ground surface at the Waste Isolation Pilot Plant (WIPP) in Carlsbad, New Mexico [30].

Low-Level Waste LLW has low radionuclide content. Similar to ILW, IAEA [45, 52] suggests that LLW be further divided on the basis of whether it consists predominantly of short-lived or long-lived radionuclides. LLW tends to be defined by what it is not (i.e., not HLW, ILW, or TRU) rather than what it is, so it can include a broad range of materials and radioactivity levels [30, 51]. For example, LLW may contain small amounts of radioactivity spread

through a large volume of material or it may contain sufficiently high levels of radioactivity to require shielding for its safe handling [15, 30, 53]. LLW wastes generated during nuclear decommissioning may involve a wide range of materials including rags, papers, filters, ion exchange resins, discarded protective clothing, contaminated soils and construction rubble, piping, and tanks.

Because of the generally low levels of radioactivity, LLW is typically disposed using near-surface burial. Depending on its physical and chemical properties, LLW may be packaged in drums, casks, special boxes, or other sealed containers [30]. Contaminated soil and construction debris may be disposed directly into the cell without a container. Some large components such as pipes and tanks may be cut up or flattened to reduce the volume.

LLW disposal facilities may be either commercial or government operations, although commercial facilities are still typically governed by government regulation. Similar to municipal landfills, LLW disposal cell designs may include a liner, and an engineered cover system may be installed to reduce water infiltration into the underlying groundwater system [15]. Monitoring systems and institutional controls are installed to ensure waste isolation (air, water, and soil) and to limit access to the disposal facility [51].

Hazardous and Mixed Wastes Depending on their age and size, decommissioning of nuclear facilities may involve management and disposal of hazardous wastes that result from the processes employed during facility operations, or from building standards used during the initial construction of the plant. For example, hazardous chemicals such as acidic, alkaline, and organic solutions used during uranium milling or fuel reprocessing may require special handling, treatment, and segregation prior to disposal. In addition, hazardous materials such as asbestos may be encountered when older buildings are dismantled and may require special treatment and disposal. If these hazardous wastes are free from radioactive contamination, their ultimate disposal path would be based on hazardous waste regulations.

Radioactive wastes may also be mixed with hazardous wastes. The management and disposal of these

wastes can be complicated, as there may be more than one agency with jurisdiction and more than one set of waste handling criteria may apply. For example, in the United States, the U.S. Environmental Protection Agency (USEPA) has authority over hazardous waste through the Resource Conservation and Recovery Act, and the USNRC and USDOE have regulatory authority over radioactive wastes through the Atomic Energy Act. In this case, the management and disposal options to be considered for decommissioning wastes may need to comply with both hazardous and radiation safety requirements [54].

Uncontaminated Wastes As with other industrial operations, nuclear facilities may also produce wastes that contain little or no radioactive contamination. If the radioactivity of these wastes falls below levels established by the applicable regulations and statutes, they may require no additional nuclear regulatory control. Also, as described in “[Safety and Performance Assessment](#)” section, some decommissioning wastes can be decontaminated below the applicable “clearance” levels and released from regulatory control [55]. IAEA [45] identifies these as “exempt” wastes and notes that they can be disposed of using conventional methods and systems.

As structures are dismantled and steel, concrete, and other surfaces (e.g., parking lots) are removed, large volumes of construction debris may be generated during nuclear facility decommissioning. Because some of these structures are not associated with nuclear-related activities, construction debris of this nature may meet the requirements of exempt waste. The specific levels of radioactivity that establish the criteria to be used in identifying exempt wastes will typically differ from country to country. As a result, the volume of exempt waste and disposal paths will also vary.

General Methodologies for Decommissioning

Specific decommissioning activities and technologies, as well as the sequence of their application, will vary depending on size, age, operational conditions, and design of a nuclear facility, and on the regulatory framework that governs decommissioning. As indicated previously, this article provides a general discussion of typical decommissioning activities.

This discussion is not a recommendation of particular approaches or technologies, and the reader is referred to the references identified in the Bibliography for further reading.

Developing a Site-Specific Decommissioning Plan

Nuclear decommissioning is perhaps most effective when the process is laid out in a decommissioning plan. For the lead organization (either commercial or government) with responsibility for decommissioning, the plan provides the opportunity to develop a strategy that, among other things, identifies specific decommissioning issues at a given site, determines the types of processes and methodologies to be used, specifies the desired end state of the facility, and establishes the schedules and financing mechanisms for decommissioning. For large, complex commercial nuclear fuel cycle facilities such as a power plant or a fuel enrichment plant, the decommissioning plan is an extensive document that is supported by a large number of underlying technical and policy reports and references. Although the specific contents of a decommissioning plan can depend on the type of facility, the regulatory framework, and other policy issues, some of the general topics to be covered include [18]:

- Facility Description and Operational History
- Radiological Status
- Alternate Decommissioning Strategies and Selection of a Preferred Alternative
- Project Management
- Decommissioning Activities
- Surveillance and Maintenance During Decommissioning
- Waste Management
- Cost Estimate and Funding Mechanisms
- Safety and Performance Assessment
- Environmental Impact Assessment
- Health and Safety for Workers and the Public
- Quality Assurance
- Emergency Planning
- Physical Security and Safeguards
- Final Radiation Survey

Typically, a site-specific decommissioning plan is developed and refined in stages. An initial plan describing major structures, systems, and features is developed

as the facility is designed and constructed. The initial plan is designed to provide basic information and establish project baselines prior to facility startup. The initial decommissioning plan is also intended to provide a framework for cost estimates (see [“Estimates of Decommissioning Costs”](#)) to ensure that the necessary funds will be in place to cover decommissioning when the facility ceases operation. During the operational life of the facility, the initial decommissioning plan is periodically updated to reflect the operator’s experience and understanding of the site. As with any large industrial facility, nuclear facilities may change as technology or regulatory oversight develops, or as the economics of the plant change. As the plant approaches the end of its operational life, the decommissioning plan becomes more detailed. The final decommissioning plan is developed just prior to a facility ceasing operations. The regulatory agency typically reviews this plan and must approve it before the operator can implement the decommissioning strategy [18]. Once the decommissioning plan is approved, it then becomes the basis for subsequent activities, although it may continue to be revised throughout a decommissioning process that can extend over decades.

Site Characterization

Site characterization provides the context for nuclear decommissioning. Ideally, site characterization includes a description of the size and location of the facility, buildings and systems, and the operational history of the site, as well as a description of spills or other releases that may affect the decommissioning process. Nonradiological hazardous process chemicals and other materials like asbestos that require special treatment and disposal may also be identified in the site characterization survey.

One objective in undertaking site characterization is to establish the preoperational baselines and background values that may be used to determine criteria for successful decommissioning. For example, an understanding of background water quality is used to establish site-specific levels for groundwater restoration, as well as action levels for monitoring. Site characterization should identify the geographical and geological context of the facility in relationship to important resources for the area such as critical habitat

or historical and cultural areas. In addition, site characterization may include a discussion of the socioeconomic impacts of the facility, because this may be an important consideration in selecting among different decommissioning strategies [1]. The site characterization and radiological surveys are intended to be of sufficient detail to provide data for planning the decommissioning effort, including selection of specific remediation techniques, establishing decommissioning schedules, estimating costs and waste volumes, and identifying important health and safety considerations to be considered during decommissioning [20].

For the purposes of nuclear decommissioning, site characterization pays special attention to the radiological status of the site, focusing on establishing the extent to which buildings, systems, equipment, soils, and water may contain residual activity [18]. These radiological “hot spots” can be determined using historical information (e.g., location of historic spills, known storage locations, sites identified by ongoing monitoring during operations), conducting surveys with radiation detection equipment, and collecting soil and water samples for subsequent analysis [18, 20, 56, 57].

A key component of site characterization is locating and maintaining existing records [58, 59]. A lack of information on past activities has been a special challenge for decommissioning Cold War legacy sites.

Selecting a Decommissioning Strategy

The reasons for taking a nuclear facility out of service may be based on economics, national policy decisions about the suitability of nuclear power, safety, or obsolescent technology [60]. As described in [“Introduction”](#) section, decommissioning strategies for nuclear facilities generally fall into one of three categories:

- A strategy of immediate decontamination and dismantlement [1, 7, 61, 62] (also defined as DECON [14, 15]) begins soon after the nuclear facility closes. For nuclear reactors, spent nuclear fuel is removed, stored, and cooled, pending permanent disposal or reprocessing, and equipment, buildings, structures, and portions of the facility that contain radioactive contaminants are either removed or decontaminated to meet regulatory requirements for releasing the property. In general, this strategy

imposes the largest requirements for resources (funding) and personnel in the short term. It takes advantage of the existence of a trained workforce with experience in operating the facility.

- Safe storage [1, 61], deferred or delayed decontamination [7, 62], or SAFSTOR [14, 15], refer to decommissioning strategies where a nuclear facility is left intact after closing, placed in a stable condition, and maintained and monitored until subsequent dismantlement and decommissioning. One purpose in choosing this strategy is to allow radioactivity to decay during a period of safe storage, perhaps on the order of decades, potentially reducing the radiological hazards and the quantity of nuclear waste that must be disposed. This strategy may also benefit from continuing developments in decommissioning technology and waste management options.

This approach places a premium on knowledge management, as the operations workforce may not be available when decommissioning begins years after the end of operations. This decommissioning strategy may be the only option if there are insufficient funds available to cover the costs of immediate decontamination and dismantlement, or if some aspects of the regulatory framework, such as a spent nuclear fuel disposal site, are not available when operations cease [7, 63].

Safe storage may allow one part of a large facility to be closed while the rest of the facility completes its life cycle. For example, one reactor unit at a nuclear power plant may be closed and the fuel removed while the remaining reactor units continue to produce electricity. When the decision is made to close the remaining portions, all of the components can then be decommissioned together, with accompanying potential benefits from optimizing the use of staff and specialized equipment for decommissioning. An example of this approach is the Peach Bottom Unit 1 reactor in York County, Pennsylvania, which was shut down in 1974 and placed in SAFSTOR. Reactor Units 2 and 3 continue to operate and are scheduled for shutdown in 2034, at which point final decommissioning will begin [64].

- A strategy of entombment [1, 7, 61, 62] or ENTOMB [14, 15] involves encasing radioactive materials onsite in a long-lived, structurally sound

material such as concrete. The entombment structures are maintained and monitored as appropriate, with institutional controls (e.g., fencing, security personnel) to limit access. For some facilities, the intent of entombment is permanent encapsulation [7], and computer models are used to simulate performance for thousands of years [19]. Because entombment effectively creates a surface waste disposal site, it is not generally a suitable decommissioning strategy for facilities associated with fuel enrichment, fuel fabrication, and fuel reprocessing [62]. In addition, an entombment strategy may also limit the options for releasing the site for reuse.

The first two decommissioning strategies may also be combined. For example, some facilities at a site may be immediately dismantled while other structures are placed in safe storage. For large sites like fuel enrichment plants, nuclear decommissioning activities may be occurring at some facilities at the same time as active operations [4]. Generally, it is anticipated that decommissioning activities will be completed in a period of decades from the end of operations [14].

Because of the wide variety of nuclear facilities, there is no unique approach to decontamination and decommissioning. Several factors may be considered in selecting the decommissioning strategy [1, 7, 62] for a specific site, including:

- The status of the policies and regulatory framework that establish, among other things, the national direction of the nuclear industry and legal requirements for nuclear decommissioning
- The financial costs (both direct and indirect) associated with a given decommissioning strategy and the amount of funding available
- The availability of waste management and disposal facilities for the types and volumes of waste to be generated during decommissioning
- Risks to health and safety of both workers and members of the public
- Potential environmental impacts associated with a given decommissioning strategy
- Knowledge management concerns and the availability of trained and experienced personnel to conduct the decommissioning activities

- The desired end state for the site, and potential socioeconomic impacts to local communities and other stakeholders, including options for release and reuse of the site after decommissioning is complete

These factors need to be considered within the context of the specific site before selecting the preferred alternative. As described previously, the reasons for selecting one alternative as opposed to another should be discussed in the decommissioning plan.

Estimates of Decommissioning Costs

The facility owner is generally responsible for ensuring that there are sufficient resources to cover activities associated with decommissioning a nuclear facility. These activities may include decontamination, decommissioning, reclamation, and groundwater restoration, as well as surveillance and monitoring that may be necessary for long-term stewardship. Cleanup of government facilities and legacy sites that were established prior to the development of a regulatory framework for nuclear decommissioning are generally the responsibility of the national (or state/provincial) governments. Funds for these government responsibilities may be raised by general appropriations (taxes) or user fees imposed on the beneficiaries of nuclear power.

Methods used to accumulate and manage funds for decommissioning commercial nuclear facilities vary from country to country [1]. One common method to establish a decommissioning fund is to impose a requirement that a portion of business revenues be set aside for decommissioning and waste management. Typically, the types of financial mechanisms that are acceptable for estimating, creating, and maintaining a decommissioning fund are either established by the regulatory agency responsible for the facility license or directly by legislation [1, 21].

In countries with a robust regulatory framework, the owner/operator is commonly required to present estimates for the cost of decommissioning activities as part of the license application. A detailed discussion of estimating decommissioning costs is contained in USNRC [21]. In general, cost estimates are specific to the size, type, and location of the facility, and they incorporate assumptions about:

- The transition between operations and facility shutdown, and the work associated with that process, such as postoperational cleanout
- The definition for the end state of the decommissioning process (e.g., unrestricted release, restricted release)
- The availability and suitability of established approaches to decommissioning methods versus the need for unique and perhaps untested technologies
- Availability and capacity of facilities for managing or disposing of residual spent fuel and radioactive waste

Once the regulatory agency or governing body evaluates and accepts the proposed cost estimates, the type and amount of the financial surety (e.g., letter of credit, prepaid cash, government bond) is established and administered in accordance with government regulations. The regulatory agency (as in Canada, United States, and Sweden) or a waste management body (as in Belgium and Spain) then reviews the fund on a regular basis, generally between 1 and 5 years [1, 21]. As identified during the review, the decommissioning fund may be updated and the amount adjusted either upwards or downwards to account for inflation, changes in technology, or completed decommissioning activities.

As noted previously, the actual costs of nuclear decommissioning can vary substantially. The IAEA provided estimates for median decommissioning costs associated with various types of nuclear facilities based on a combination of expert judgment and decommissioning experience and using assumptions with respect to operating life and time to decommission [4]. These estimates (in US \$, 2003 value) are summarized in Table 2. Because the cost of decommissioning will depend strongly on site-specific issues such as local geology, facility age and design, and operational history, the actual costs for a given site may fall outside these ranges.

The estimates reported in IAEA [4] are general in nature and are not intended to bound all potential costs, particularly for very large government facilities. One general conclusion that can be made from these estimates is that management of spent nuclear fuel and radioactive wastes can represent a significant

Nuclear Facilities, Decommissioning of. Table 2 Estimated costs associated with decommissioning different nuclear facilities [4]

Type of nuclear facility	Estimated decommissioning cost (min to max [median] in US \$ million, 2003 value)	Operational life (years)	Time to decommission (years)
Uranium milling	0.800	25	1
Uranium conversion/recovery	150	30	3
Uranium enrichment	600	30	10
Fuel fabrication	250	30	2
Nuclear power reactor ^a	250–500 (350)	40	10 (after 5-year transition period)
Fuel reprocessing	800	30	15
Industrial facilities	0.050–3 (0.200)	20	1

^aCost estimates for decommissioning nuclear power reactors do not include the processing of operational waste, removal and disposition of spent nuclear fuel, the draining of operational systems, or the development of a waste disposal facility

proportion of the total costs of decommissioning. For example, estimated costs for decommissioning the Sequoyah Fuels uranium conversion site near Gore, Oklahoma, varied from US \$19 million for the no-action alternative of long-term stewardship, to about US \$36 million for onsite disposal of most contaminated wastes, to as much as US \$254 million for transportation and offsite disposal of all contaminated wastes [25]. These cost issues indicate the importance of accurately characterizing a site and identifying appropriate opportunities to decontaminate, recycle, and reuse materials.

Safety and Performance Assessment

Safety is among the highest priority issues in nuclear decommissioning. Safety assessments, typically developed by both operators and regulators, are engineering analyses that involve calculations and computer simulations to evaluate potential radiological doses. The purpose of the safety assessment is to identify and evaluate potential hazards to ensure that nuclear decommissioning can be done in a manner that is safe for workers, members of the public, and the environment [65].

In general, the safety assessment should be systematic and be linked to relevant safety criteria, taking into consideration potential radiological doses to workers

and members of the public, discharges to the environment, and exposure to chemical and other nonradiological hazards [18, 57, 65, 66]. To meet these objectives, a safety assessment consists of:

- Estimates of system performance for all the situations selected
- Evaluation of the level of confidence in the estimated performance
- Overall assessment of compliance with safety requirements

The standards and criteria to be used in developing these assessments vary from country to country depending on the regulatory framework that is in place. Also, the nature of the safety assessment may vary depending on the complexity of the decommissioning strategy needed for a given nuclear facility. A general framework proposed by the IAEA [64] includes:

- The scope of the assessment, based on the physical state of the nuclear facility
- The objectives of the assessment
- The applicable safety requirements and criteria to be used in evaluating potential exposures to workers, members of the public, and the environment
- Outputs from the safety assessment, generally in the form of doses that can be compared to the relevant safety requirements and criteria

- A description of the approach used to implement the safety assessment, whether through simplified calculations or complicated computer models, and a discussion of how the approach is appropriate to the magnitude and time frames of the potential hazards
- Time frames for all phases of the decommissioning activities considered in the safety assessment
- A definition of all phases of nuclear decommissioning and the anticipated end points for each phase
- A definition of the final end state of the facility that is anticipated after all decommissioning activities are complete

In practice, a safety assessment starts with a description of the facility and all of the anticipated decommissioning activities that comprise the decommissioning strategy (see [“Selecting a Decommissioning Strategy”](#)). To evaluate off-normal scenarios, a safety assessment also includes identification of potential hazards and initiating events (both natural events such as earthquakes and human-made events such as fire), and potential exposure pathways. These are screened, usually on the basis of probability of their occurrence and the resulting consequence should they occur. Plausible scenarios are developed and used in the engineering analysis to quantify the likelihood and magnitude of potential radiological and safety consequences. In addition, the safety assessment should identify relevant experience and lessons learned from the decommissioning of similar facilities, if available.

For disposing of longer lived radionuclides, some regulatory frameworks and decommissioning strategies use a performance assessment that provides a quantitative evaluation of potential releases of long-lived radionuclides over time periods of hundreds to thousands of years or longer. Similar to safety assessments, but with a much longer time horizon, performance assessments typically involve computer simulations based on site-specific features, events, and processes (biological, physical, and chemical) that may affect the long-term performance of engineered barrier systems. A performance assessment also simulates the release of radionuclides from any engineered barrier system and their subsequent migration through the geosphere surrounding the facility. Finally, future

potential radiological doses may be calculated for a hypothetical receptor group located away from the facility [38]. Because of the long time frame, it is not possible to include all potential conditions that might affect performance, so simplified models, or abstractions, are used to simulate important aspects of the engineered and geological systems. This can introduce uncertainty into the calculations that should be characterized and evaluated to determine whether there is sufficient confidence that the applicable regulatory criteria will be met [19, 38].

In essence, the idea of safety and performance assessments is intended to answer these questions: What can go wrong? How likely is it? What are the consequences? [38]. Where there is uncertainty in conceptual models of the system or in the model parameters, simplifying assumptions should be chosen in such a way that ensures the models will be transparent, conservative, and not underestimate potential radiological doses. Statistical analysis and sensitivity studies can help to characterize the nature and relative importance of the uncertainty. Depending on the results of the safety and/or performance assessment, the planned decommissioning activities may be modified to reduce risks, or compensate for uncertainty or limited information [66].

Decontaminating Structures, Equipment, and Components for Reuse or Recycling

The objectives of decontamination include reducing the potential radiological exposure of workers and members of the public during decommissioning; minimizing the volume of radioactive waste; and increasing the potential for reusing and recycling equipment, material, and land at the site of a nuclear facility [67].

Although specific technologies to be used during decommissioning activities will vary depending on size, age, operational conditions, and design, general decontamination methods include [68–70]:

- Chemical methods that use agents such as acids, oxidants, or chemical foams and gels to remove contamination fixed to surfaces. The nature of the chemical agent may be determined based on the properties of the surface to be decontaminated. Chemical decontamination methods may generate a liquid waste stream that requires further

treatment to remove radioactive wastes (and generate subsequent secondary waste streams).

- Mechanical methods that rely on cutting, grinding, or other physical techniques to remove contaminated surfaces or layers.

The techniques used for decontamination will vary by material and from site to site, depending on issues such as the operational history of the facility, the level of contamination, and the type of material to be decontaminated [67, 69]. For example, chemical decontamination methods may not be appropriate for porous materials such as concrete, where fluids may migrate into the material.

In addition to technical feasibility, other considerations in selecting decontamination techniques may include the applicable clearance criteria and the potential doses to workers or members of the public to reach these levels. The selection process may also include evaluating the cost of decontamination versus the cost of disposal without decontamination, taking into account estimates of the volume, nature, category, and activity of any primary and secondary wastes that might result. For example, decontamination methods should be selected in such a way as to minimize the amounts of secondary wastes (e.g., cutting fluids or washdown fluids) produced. It is also important to ensure that selected decontamination methods are compatible with and do not compromise existing or planned key treatment, conditioning, storage, and disposal systems [67]. For example, high-pressure jets may not be an appropriate method for cleaning liquid waste storage tanks that are decades old and have experienced some corrosion.

Dismantling and Removing Buildings, Structures, and Equipment from the Site

Equipment removed from buildings (Fig. 4) during the cleanup process is typically categorized as (1) salable for unrestricted use after radiation checks and decontamination; (2) potentially contaminated, but may be salvageable for sale to other nuclear facilities; and (3) contaminated and must be disposed. The responsible regulatory authority generally establishes the levels of contamination for these different categories.

Dismantling and removing buildings, foundations, structures, and equipment from a nuclear facility is

sometimes referred to as “construction in reverse,” and similar types of heavy equipment (e.g., trucks, cranes, earthmoving equipment) are used during dismantling (see Figs. 4 and 5). In addition, cutting equipment and even demolition explosives may be needed to break large structures down into smaller pieces that can be handled and transported more easily (Figs. 5 and 6). For high-radiation areas, remote techniques such as underwater cutting may be necessary to provide for worker radiation protection [67].

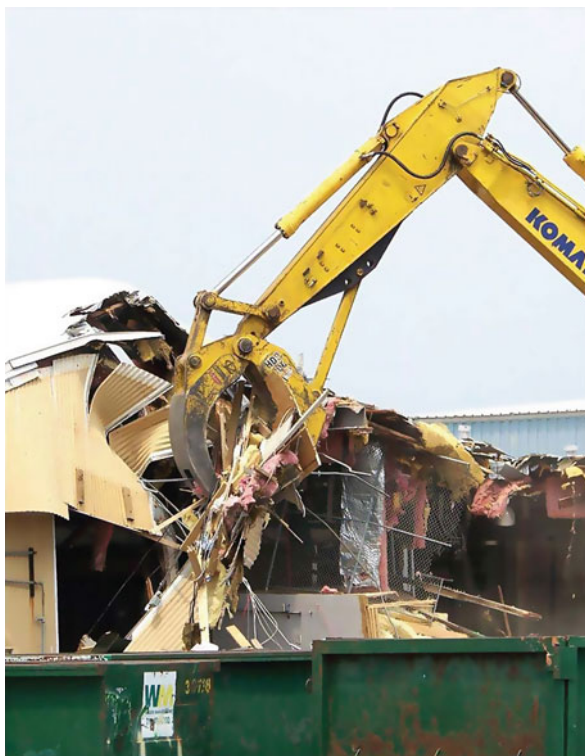
Where salvage is not economically or technically feasible, larger pieces of equipment (e.g., pipes, tanks) may be cut up and flattened, and building materials reduced to rubble to minimize waste volume and make handling for subsequent disposal easier. How much volume needs to be reduced may depend on the



Nuclear Facilities, Decommissioning of. Figure 4
A steam generator is removed from a reactor building [76]



Nuclear Facilities, Decommissioning of. Figure 5
Demolition of a reactor containment building [76]



Nuclear Facilities, Decommissioning of. Figure 6
Dismantling industrial buildings at a nuclear facility [42]

available waste disposal capacity and regulations that govern transportation and disposal. Pavement in roads and parking lots may also be removed for disposal as construction debris.

Remediating Contaminated Soils and Groundwater

Remediating residual radioactivity in soils and water generally uses technologies that are well established in the environmental industry. Evaluating and selecting a specific technique is an important part of decommissioning and depends on many site-specific issues such as the physical and chemical state of the residual radioactive material, the availability of equipment and a trained workforce, the availability of appropriate waste disposal capacity, and stakeholder sentiment. The following discussion is intended only as a brief summary of commonly available methods for remediation.

The most common remediation method used for contaminated soils involves excavating the soil for disposal in a LLW facility (Fig. 7). If the contamination layer



Nuclear Facilities, Decommissioning of. Figure 7
Removing uncontaminated soil and debris for offsite disposal [42]

is not deep, this may be an economically viable approach. If the contamination is widespread, however, the volume of material to be disposed in this fashion may be quite large and may increase the costs of waste disposal. Alternatively, the soil may be excavated and treated using physical and chemical separation techniques to isolate and reduce the volume of contaminated material [27]. For example, in some cases, residual radioactivity may be more closely associated with fine clay-sized particles in the soil that can be separated by using screens or other physical separation methods. Chemical extraction methods use solutions such as organic acids to bind to the contamination and remove it from the soil. After separation of the contaminated and uncontaminated soil fractions, the uncontaminated soils are used as clean fill, and contaminated soils are treated further or processed for disposal. If chemical methods are used, the leachate may be treated to remove the dissolved radionuclides to meet water standards.

Contaminated surface water and groundwater typically require treatment to meet applicable standards. Widely available technologies for contaminated groundwater include pumping and treating to pull the contaminant plume back toward the extraction well. The contaminated water is then treated through a process such as ion exchange and is either injected back into the aquifer or discharged to a suitable surface water disposal system. A more recent technology for treating contaminated groundwater is the construction of a passive permeable reactive barrier or slurry wall system (Fig. 8). Built below the ground surface so that they intercept the groundwater plume, the reactive materials in the barrier are selected to chemically react with the contaminants and immobilize them in the subsurface. Reactive materials that may be used to remove uranium from groundwater include different forms of iron, such as metallic (zero-valent) or amorphous ferrihydrite [27]. The contamination will remain in place until the barrier is excavated, and barrier

longevity and long-term performance are important engineering issues.

Other technologies rely on in place or in situ methods to reduce or immobilize soil and water contamination. For soils, contamination may be reduced or immobilized through mixing in amendments such as apatite or phosphate that will chemically react with the contamination. Alternatively, soil contamination may be immobilized in place by grouting or capping. Bioremediation is a collection of more recent technologies that use biological organisms to preferentially extract or otherwise break down toxic and radiological contamination from both soils and groundwater [71]. For example, sunflowers have been demonstrated to take up uranium from waste at a site in Ashtabula, Ohio, and at a small pond contaminated with uranium near the Chernobyl nuclear power plant site in Pripyat, Ukraine [27, 71].

In addition to active remediation technologies, monitored natural attenuation of contaminated soils and groundwater may be also be applicable, if it can be demonstrated to meet applicable criteria in a reasonable timeframe [72]. This method relies on monitoring soils and groundwater while the natural physical, chemical, or biological processes already occurring at the site contain and reduce volume, mass, and toxicity of the contamination in place. Where site conditions such as soil type and groundwater flow are favorable, monitored natural attenuation can be an attractive option because it is typically less disruptive and costly than more active remediation measures. However, because it can be perceived by stakeholders as “doing nothing,” monitored natural attenuation is generally proposed as one part of a broader remediation strategy and combined with active remediation measures [72]. For example, a contamination source may be excavated and removed while monitored natural attenuation is implemented for the associated groundwater contaminant plume.



Nuclear Facilities, Decommissioning of. Figure 8
Building a slurry wall to remediate contaminated groundwater [42]

Waste Management and Disposal

As described in “Waste Management and Disposal” section, decommissioning wastes can fall into several different classes. Waste classification ultimately depends on the type of historical operations at the site and will influence the decommissioning strategy

and waste management options that are selected for a facility. Once the physical, chemical, and radiological properties are characterized and the waste is classified and the applicable criteria are determined, the basic options for waste management are either onsite storage/disposal or transport offsite to an approved waste disposal facility.

For nuclear power plants, one of the final phases of operations before decommissioning is to remove the fuel from the reactor (defuel) and place the spent nuclear fuel in interim storage, either in pools or dry casks. This step would be necessary regardless of the decommissioning strategy selected [69]. As they are generated during operations, other wastes may be collected, segregated, chemically adjusted, and decontaminated onsite, and then placed in temporary monitored storage onsite until a final disposition path is determined [44]. A partial list of examples of the types of wastes that may be encountered during the decommissioning of nuclear facilities is presented in Table 3.

Waste may be treated to prepare it for final disposal. Treatment concepts include volume reduction and separation and removal of radionuclides and other hazardous wastes [44]. Some treatment options include

Nuclear Facilities, Decommissioning of. Table 3 Partial list of the types of waste materials for different nuclear facilities

Type of nuclear facility	Examples of waste materials
Uranium mine	Waste rock
	Fuels and lubricants
	Contaminated soils and groundwaters
Uranium mill	Drums
	Insoluble waste and filter materials
	Liquid effluent
	Tailings and sludges
	Liquid nitrates
	Ion exchange resins
	Tanks, pipes, and equipment
	Contaminated soils and groundwater

Nuclear Facilities, Decommissioning of. Table 3 (Continued)

Type of nuclear facility	Examples of waste materials
Uranium conversion	Solid CaF_2
	CaF_2 sludges with/without minor uranium
	Non-radiological chemical waste
	Tanks, pipes, and equipment
	Contaminated soils and groundwaters
Uranium enrichment	Depleted uranium tails
	Tanks, pipes, and equipment
	Contaminated soils and groundwaters
Fuel fabrication	Uranium scrap material
	Filters
	Wash water and decontamination/cleaning solutions
	Waste oils
	Spent acids and solvents
	Equipment
	Contaminated soils and groundwaters
Nuclear power plants	Reactor vessel and internal components
	Coolant system equipment and components
	Activated concretes and steels
	Evaporator concentrates
	Tanks, pipes, and equipment
	Contaminated debris and soils
Spent fuel reprocessing	Filters
	Activated and contaminated metal components
	Spent solvents, decontamination and metal cleaning agents
	Fuel cladding
	Laboratory analytical equipment and solutions
	Tanks, pipes, and equipment
	Contaminated debris, soils, and groundwater

incineration or compaction to reduce volume, and evaporation or ion exchange to remove radionuclides from liquid wastes. Some of these techniques can generate secondary wastes such as liquids, sludges, and filters that need to be managed as well.

Radioactive wastes may also be conditioned to produce a form that is more suited for handling, transportation, storage, and disposal. Low-level and intermediate-level wastes may be immobilized by mixing with grouts, cements, and bitumen, while liquid high-level wastes from fuel reprocessing may be vitrified into a glass waste form or otherwise modified to produce a solid waste form. These wastes are placed in packages or specially designed containers for interim storage, transportation, and subsequent offsite disposal (Fig. 9). This may be at a licensed disposal site, a specially constructed disposal cell, or an existing tailings impoundment that is being reclaimed and used as a disposal cell. The materials must meet waste acceptance criteria for the disposal facility.



Nuclear Facilities, Decommissioning of. Figure 9
Low-level radioactive waste packages for offsite disposal [42]

Alternatively, waste may be disposed in place with monitoring and an engineered system for long-term disposal [42]. Onsite waste management, including a decommissioning strategy of entombment, must comply with applicable regulations for near-surface storage and disposal facilities [62]. In this case, a performance assessment (see “[Safety and Performance Assessment](#)”) may be used to evaluate both engineered and natural barrier performance over long times. Depending on the nature of the waste and the site-specific conditions, different design options (backfill, concrete vaults, engineered covers) may be evaluated along with a combination of long-term monitoring and institutional controls [44]. Institutional controls such as fencing, signage, and physical security are generally assumed to be effective only for a period on the order of hundreds of years, while the performance of engineered and natural barriers is evaluated over periods as long as 10,000 years [19, 20].

Conducting Final Inspections and Surveys

During decommissioning, the national regulatory agency will typically conduct onsite reviews and inspections to ensure that the approved decommissioning plan is being followed [19] and the decommissioning activities are being conducted in a safe manner that complies with applicable regulations. After decommissioning is complete, there should be a final survey to evaluate the residual radioactivity that remains at the site. The specific goal of this survey is to determine the extent to which the site complies with the criteria set by the governing regulatory authority for subsequent reuse and/or release of the site [57, 67] – one of the objectives of decommissioning (see “[Definition of the Subject](#)”).

This survey may be carried out in phases, as decommissioning work is completed, to enable parts of the site to be released from regulatory control. The final survey data are submitted to the regulatory authority, including a description of the applicable reuse/release criteria, the methods and procedures used, and the measurement results. Typically, the sampling and surveys will be focused on areas such as previous waste burial sites or spill locations identified as contaminated during the site characterization survey. These areas are sometimes called Class 1 areas [57],

and sampling and surveys will be more extensive. Areas that are less likely to be contaminated, called Class 2 and Class 3 areas, receive lesser surveying and sample coverage. The results typically include an analysis to demonstrate the statistical significance of the results as compared to natural background radiation levels [20, 56, 57, 67].

Reclaiming Disturbed Land

After decommissioning and dismantlement of buildings and structures is complete, the decommissioning strategy may call for reclaiming disturbed land to preoperational condition. This is generally accomplished by adding clean topsoil, installing drainage and erosion controls if necessary, regrading and recontouring the surface to match the surrounding topography, and revegetating with native vegetation (Fig. 10). In addition, land reclamation efforts should include monitoring to ensure that drainage and erosion controls are functioning as intended, revegetated areas

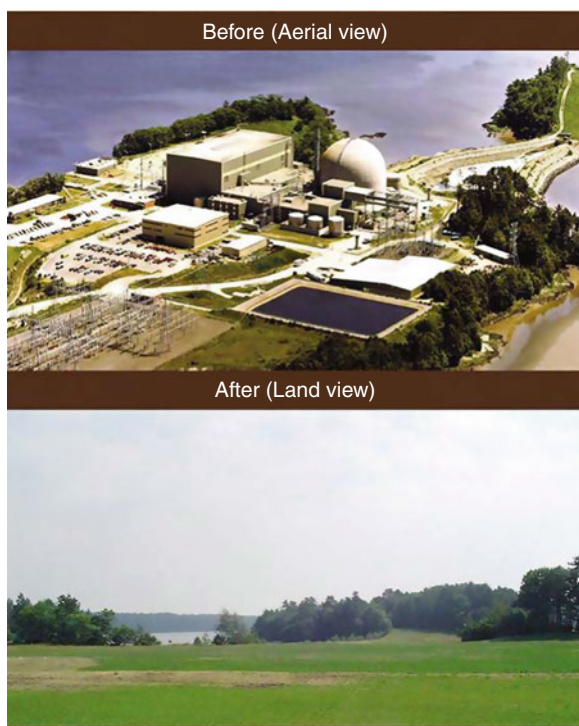
are stable, and invasive species have not become established. Many of these techniques are similar to those used to reclaim surface mines [73].

The amount of disturbed land to potentially reclaim can vary significantly for different nuclear facilities, especially conventional uranium mining and milling operations that produce significant quantities of waste rock and tailings covering tens to hundreds of hectares (acres). The type of reclamation will depend, in part, on the intended use of the land once the license is terminated. For example, a site intended for unrestricted release as a recreation area may have a more extensive land reclamation program than a site that will ultimately be released for future industrial uses.

Status of Decontamination and Decommissioning

Because commercial nuclear fuel cycle facilities and nuclear power plants are governed at the national level using different policy, statutory, and regulatory frameworks, the status of nuclear decommissioning around the world is difficult to estimate [4]. A brief summary based on the estimates of international nuclear organizations such as the IAEA, NEA, and WNA is provided here and in Table 4.

- The leading countries for uranium production include Canada, Kazakhstan, and Australia, followed by Namibia, Russia, Niger, Uzbekistan, and the United States [40]. Uranium has been mined for decades, and some of the older mines are small and not well documented. The WNA estimates that more than 100 mines have been “retired from operations,” although decommissioning is not complete at all of them [5].
- Uranium mills exist in countries with known uranium ore reserves. As of August 2003, the IAEA reported that 294 uranium mills were operating worldwide, with 231 plants shut down or in decommissioning and 149 fully decommissioned. Eight additional uranium mills were under construction [4].
- Uranium conversion plants to produce UF_6 operate in the United States, Canada, France, United Kingdom, China, and Russia. As of August 2003, the IAEA reports that 29 uranium conversion facilities



Nuclear Facilities, Decommissioning of. Figure 10 Reclamation and revegetation of site after decommissioning a commercial nuclear power plant [77]

Nuclear Facilities, Decommissioning of. Table 4 Decommissioning status for nuclear fuel cycle facilities as of August 2003 ([4], Table 3)

Type of nuclear facility	Operating	Under construction	Shut down/being commissioned	Decommissioned
Uranium milling	294	8	231	149
Uranium conversion/recovery	29	1	14	2
Uranium enrichment	21	2	7	5
Fuel fabrication/heavy water production	66	5	27	23
Fuel reprocessing	13	3	18	13

were operating worldwide, with 14 plants shut down or in decommissioning and 2 fully decommissioned. One additional uranium conversion facility was under construction [4].

- Major enrichment plants are operated in the United States, France, and Russia, with smaller plants located in the United Kingdom, Netherlands, Germany, Japan, and China [28]. As of August 2003, the IAEA reports that 21 fuel enrichment facilities were operating worldwide, with 7 plants shut down or in decommissioning and 5 fully decommissioned. Two additional fuel enrichment plants were under construction [4].
- Fuel fabrication and/or heavy water production facilities operate in most countries with nuclear programs. As of August 2003, the IAEA reports that 66 fuel enrichment and heavy water production facilities were operating worldwide, with 27 plants shut down or in decommissioning and 23 fully decommissioned. Five additional fuel enrichment plants were under construction [4].
- At present, 436 commercial nuclear reactors are operating worldwide in many countries in North America, Europe, and Asia [9]. A total of 112 nuclear reactors have been placed in long-term shut down, and 14 have been completely decommissioned [4, 9]. Russia has the largest number of research and test reactors, followed by the United States, Japan, France, Germany, and China. In addition, research and test reactors are also located in developing countries in Africa, South America, and Asia. After peaking in the mid-1970s at about 370 reactors in 55 countries, the number of

research and test reactors worldwide has declined sharply. As of August 2003, the IAEA reported that there were about 287 research and test reactors and critical assemblies in operation worldwide, with 8 more under construction. A total of 214 research and test reactors were shut down or being decommissioned, and 173 were reported as being decommissioned [4].

- As of August 2003, the IAEA reports that 13 fuel reprocessing plants were operating worldwide, with 18 plants shut down or in decommissioning and 13 fully decommissioned. Three additional fuel reprocessing plants were under construction [4].

Future Directions

As noted previously, interest in commercial nuclear power has been revived as a result of concerns with carbon emissions from traditional fossil fuels. As a result, many nations have reported an increase in proposed construction projects for new reactors. Although many of these proposed projects will use current Generation II and III reactor designs, there are research and development programs focused on designing the Next Generation IV reactors [37].

The different current and future design types and sizes of reactors will make decommissioning inherently a site- and reactor-specific process. Each current and future reactor design will have its own design-specific decommissioning requirements that must be taken into consideration. In addition, changes in reactor design or the use of nuclear fuels based on MOX or

thorium may lead to different fuel requirements that will change the nature of decommissioning at the front end of the nuclear fuel cycle.

Much of the decommissioning experience gained to date is focused on an open, or “once through” uranium fuel cycle. Current programs such as the Global Nuclear Energy Partnership are evaluating the potential of a closed fuel cycle, where fuel is reprocessed and used more than once to generate power [39]. This approach, along with changes in ongoing nuclear reactor operations, would change the amount and characteristics of spent nuclear fuel for handling, storage, and disposal. In addition, closing the fuel cycle would require new nuclear fuel cycle facilities and change the waste streams produced at the back end of the fuel cycle. For these reasons, a change from an open to a closed nuclear fuel cycle may need new waste disposal options that will influence the selection of the decommissioning strategy for these types of facilities.

Although new decontamination techniques are continually being developed [27, 69], existing technology that has been previously applied with demonstrated success may continue to be preferred by stakeholders (members of the public, government agencies) to an untried method, particularly if trained staff and specialized equipment are readily available. As decommissioning costs and potential benefits are better understood, newer techniques may be applied more. In addition, many nations are looking at extending the service life of existing nuclear facilities and reactors. At the same time, the nuclear workforce is aging, placing a premium on knowledge management and recordkeeping to ensure valuable information and experience is not lost [58, 74, 75]. Decommissioning decisions are perhaps many decades in the future, but should be taken into consideration as designs are developed for advanced reactors and the nuclear facilities that support them.

Bibliography

Primary Literature

1. NEA (2002) The decommissioning and dismantling of nuclear facilities. Paris, France. Status, approaches, challenges. Report No. NEA 3714. Nuclear Energy Agency/Organisation for Economic Co-Operation and Development, Issy-les-Moulineaux. <http://www.nea.fr/html/rwm/reports/2002/3714-decommissioning.pdf>. Accessed 28 Apr 2009
2. Commission of the European Communities (Oct 2007) Communication from the commission to the council and the European parliament. Nuclear illustrative programme. Presented under Article 40 of the Euratom Treaty for the Opinion of the European Economic and Social Committee. Commission of the European Communities, Brussels
3. NEA (2009) NEA annual report 2008. Nuclear Energy Agency, Issy-les-Moulineaux. <http://www.nea.fr/html/pub/activities/ar2008/AR2008.pdf>. Accessed 20 May 2009
4. IAEA (Aug 2004) Status of the decommissioning of nuclear facilities around the world. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/Pub1201_web.pdf. Accessed 20 May 2009
5. WNA (Sept 2006) Safe decommissioning of civil nuclear industry sites. World Nuclear Association, London. http://www.world-nuclear.org/reference/position_statements/decommissioning.html. Accessed 20 May 2009
6. O'Sullivan P, Pescatore C, Tripputi I (2009) Applying decommissioning experience to the design and operation of new nuclear power plants. NEA News 27:13–15, <http://www.nea.fr/html/pub/newsletter/2009/applying-decommissioning.pdf>. Accessed 8 July 2009
7. IAEA (Nov 2005) Selection of decommissioning strategies: issues and factors. Report by an expert group. IAEA-TECDOC-1478. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/TE_1478_web.pdf. Accessed 20 May 2009
8. IAEA (2008) International atomic energy agency annual report 2007: 50 years of atoms for peace. IAEA, Vienna. http://www.iaea.org/Publications/Reports/Anrep2007/anrep2007_full.pdf. Accessed 20 May 2009
9. IAEA (2009) Latest news related to the power reactor information system and the status of nuclear power plants. IAEA, Vienna. <http://www.iaea.org/cgi-bin/db.page.pl/pris.main.htm>. Accessed 12 June 2009
10. Reisenweaver D, Laraia M (2000) Preparing for the end of the line. Radioactive residues from nuclear decommissioning. IAEA Bull 42:51–54
11. USNRC (2009) Expected new nuclear power plant applications: updated 4 Feb 2009. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reactors/new-reactors/new-licensing-files/expected-new-rx-applications.pdf>. Accessed 20 May 2009
12. IAEA (2008) Energy, electricity and nuclear power estimates for the period up to 2030. Reference data series no. 1. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/RDS1-28_web.pdf. Accessed 12 June 2009
13. USNRC (June 2009) Generic environmental impact statement for in-situ leach uranium milling facilities. Final report. NUREG-1910, vols 1 and 2. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1910/>. Accessed 5 June 2009
14. USNRC (July 2000) Regulatory guide 1.184. Decommissioning of nuclear power reactors. U.S. Nuclear Regulatory Commission, Washington, DC

15. USNRC (June 2000) Staff responses to frequently asked questions concerning decommissioning of nuclear power reactors. Final report. NUREG-1628. U.S. Nuclear Regulatory Commission, Washington, DC
16. NEA (2005) Decommissioning of nuclear power facilities. It can and has been done. Issy-les-Moulineaux, Nuclear Energy Agency. <http://www.nea.fr/html/rwm/reports/2004/nea5728-decom.pdf>. Accessed 28 May 2009
17. WNA (Jan 2009) The nuclear fuel cycle. World Nuclear Association, London. <http://www.world-nuclear.org/info/inf03.html>. Accessed 10 June 2009
18. IAEA (Nov 2005) Standard format and content for safety related decommissioning documents. Safety reports series no. 45. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/Pub1214_web.pdf. Accessed 20 May 2009
19. USNRC (Sept 2006) Consolidated decommissioning guidance. decommissioning process for materials licensees. Final report. NUREG-1757, vol 1, Rev 2. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1757/v1/sr1757v1r2.pdf>. Accessed 20 May 2009
20. USNRC (Sept 2006) Consolidated decommissioning guidance. Characterization, survey, and determination of radiological criteria. Final report. NUREG-1757, vol 2, Rev 1. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1757/v2/sr1757v2r1.pdf>. Accessed 20 May 2009
21. USNRC (Sept 2003) Consolidated NMSS decommissioning guidance. Financial assurance, recordkeeping, and timeliness. Final report. NUREG-1757, vol 3. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1757/v3/sr1757v3.pdf>. Accessed 20 May 2009
22. USDOE (Feb 1995) Decommissioning of U.S. uranium production facilities. DOE/EIA-0592. U.S. Department of Energy, Energy Information Administration, Washington, DC
23. Taylor G, Farrington V, Woods P, Ring R, Molloy R (Aug 2004) Review of environmental impacts of the acid in-situ leach uranium mining process. CSIRO land and water client report. Commonwealth Scientific and Industrial Research Organisation, Clayton South
24. IAEA (Aug 2004) The long term stabilization of uranium mill tailings. Final report of a co-ordinated research project, 2000–2004. IAEA-TECDOC-1403. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/TE_1403_web.pdf. Accessed 20 May 2009
25. USNRC (May 2008) Environmental impact statement for the reclamation of the sequoyah fuels corporation site in Gore, Oklahoma. Final report. NUREG-1888. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1888/sr1888-intro-chaptr11.pdf>. Accessed 5 June 2009
26. USNRC (June 2009) Fact sheet: Uranium enrichment. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reading-rm/doc-collections/fact-sheets/enrichment.pdf>. Accessed 5 June 2009
27. USEPA (Dec 2006) Depleted uranium technical brief. EPA-402-R-06-011. Office of Air and Radiation, U.S. Environmental Protection Agency, Washington, DC. <http://www.epa.gov/radiation/docs/402-r-06-011.pdf>. Accessed 2 July 2009
28. Bodansky D (2004) Nuclear energy. Principles, practices, and prospects, 2nd edn. Springer/LLC, New York
29. USDOE (Apr 1999) Final programmatic environmental impact statement for alternative strategies for the long-term management and use of depleted uranium hexafluoride. DOE/EIS-0269. U.S. Department of Energy, Office of Nuclear Energy, Science, and Technology, Washington, DC
30. Interstate Technology and Regulatory Council (Jan 2008) Decontamination and decommissioning of radiologically contaminated facilities. Interstate Technology and Regulatory Council, Washington, DC. <http://www.itrcweb.org/Documents/RAD5.pdf>. Accessed 5 June 2009
31. USNRC (June 2007) Uranium enrichment. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/materials/fuel-cycle-fac/ur-enrichment.html>. Accessed 5 June 2009
32. GAO (Nov 2007) Uranium enrichment: extension of decontamination and decommissioning fund may be needed to cover projected cleanup costs. GAO-08-277T. U.S. Government Accountability Office, Washington, DC. <http://www.gao.gov/new.items/d08277t.pdf>. Accessed 12 June 2009
33. IAEA (Oct 2004) Research reactors worldwide. IAEA, Vienna. http://www-naweb.iaea.org/naweb/physics/ACTIVITIES/Research_Reactors_Worldwide.htm. Accessed 14 June 2009
34. USNRC (Aug 2008) Power reactors. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reactors/power.html>. Accessed 14 June 2009
35. Lake JA, Bennett RG, Kotek JF (Jan 2003) Next generation nuclear power. new, safer, and more economical nuclear reactors could not only satisfy many of our future energy needs but could combat global warming as well. Sci Am <http://www.scientificamerican.com/article.cfm?id=next-generation-nuclear>. Accessed 14 June 2009
36. USDOE (Nov 2008) Next-generation nuclear energy. Generation IV program fact sheet. U.S. Department of Energy, Office of Nuclear Energy, Washington, DC. <http://nuclear.energy.gov/pdfFiles/factSheets/NextGenerationNuclearEnergy.pdf>. Accessed 14 June 2009
37. USDOE (Aug 2008) Next generation nuclear plant licensing strategy. A report to Congress. U.S. Department of Energy, Office of Nuclear Energy, Washington, DC. http://nuclear.energy.gov/pdfFiles/NGNP_reporttoCongress.pdf. Accessed 14 June 2009
38. USNRC (2009) Performance assessment for waste disposal and decommissioning. Updated 11 Feb 2009. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/about-nrc/regulatory/performance-assessment.html>. Accessed 15 May 2009
39. USDOE (Oct 2008) Draft global nuclear energy partnership programmatic environmental impact statement. DOE/EIS-0396.

- U.S. Department of Energy, Office of Nuclear Energy, Washington, DC. <http://www.gc.energy.gov/NEPA/1003.htm>. Accessed 8 July 2009
40. WNA (June 2009) World uranium mining. World Nuclear Association, London. <http://www.world-nuclear.org/info/inf23.html>. Accessed 5 June 2009
 41. Andrews A (27 Mar 2008) Nuclear fuel reprocessing: U.S. policy development. CRS report for Congress, Order code RS22542. Congressional Research Service, Washington, DC. <http://www.fas.org/spp/crs/nuke/RS22542.pdf>. Accessed 1 July 2009
 42. USDOE (Nov 2008) Revised draft environmental impact statement for decommissioning and/or long-term stewardship at the West Valley demonstration project and Western New York nuclear service center. DOE/EIS-0226-D. U.S. Department of Energy, West Valley
 43. IAEA (Dec 2007) Disposal aspects of low and intermediate level decommissioning waste. Results of a coordinated research project 2002–2006. IAEA-TECDOC-1572. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/TE_1572_web.pdf. Accessed 24 June 2009
 44. IAEA (Apr 2000) Radioactive waste management profiles: a compilation of data from the waste management database no. 3. IAEA, Vienna. <http://www-pub.iaea.org/MTCD/publications/PDF/rwmp-3/RWMP-V3.pdf>. Accessed 24 June 2009
 45. IAEA (Apr 1994) Classification of radioactive waste. A safety guide. Safety series no. 111-G-1.1. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/Pub950e_web.pdf. Accessed 24 June 2009
 46. Environmental Law Institute (1999) Institutional controls case study: Grand junction. Environmental Law Institute, Washington, DC. <http://www.elistore.org/Data/products/d9.14.pdf>. Accessed 23 June 2009
 47. USNRC (2003) Standard review plan for the review of a reclamation plan for mill tailings sites under title II of the Uranium Mill Tailings Radiation Control Act of 1978. Final report. NUREG-1620, Rev 1. U.S. Nuclear Regulatory Commission, Washington, DC
 48. USNRC (Nov 2008) Design, construction, and inspection of embankment retention systems at uranium recovery facilities. Regulatory guide 3.11. U.S. Nuclear Regulatory Commission, Washington, DC
 49. IAEA (May 2005) Guidebook on environmental impact assessment for *in situ* leach mining projects. IAEA-TECDOC-1428. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/TE_1428_web.pdf. Accessed 20 May 2009
 50. National Research Council (2001) Disposition of high-level waste and spent nuclear fuel. The continuing societal and technical challenges. National Academy Press, Washington, DC
 51. USDOE (Jan 2007) Radioactive waste management manual. DOE M 435.1-1. U.S. Department of Energy, Office of Environmental Management, Washington, DC. <http://www.directives.doe.gov/pdfs/doe/doetext/neword/435/m4351-1c1.pdf>. Accessed 24 June 2009
 52. IAEA (1999) Near surface disposal of radioactive waste safety requirements. IAEA safety standards series, Safety guide no. WS-R-1. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/P073_scr.pdf. Accessed 29 June 2009
 53. IAEA (2008) Managing low radioactivity material from the decommissioning of nuclear facilities. Technical reports series no. 462. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/trs462_web.pdf. Accessed 12 June 2009
 54. League of Women Voters Education Fund (1993) The nuclear waste primer. The League of Women Voters Education Fund, Washington, DC
 55. IAEA (2004) Application of the concepts of exclusion, exemption and clearance. IAEA safety standards series, Safety guide no. RS-G-1.7. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/Pub1202_web.pdf. Accessed 29 June 2009
 56. USNRC (June 2001) Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM). NUREG-1575, Rev 1/EPA 402-R-97-016, Rev 1/DOE/EH-0624, Rev 1. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.epa.gov/radiation/marssim/obtain.html>. Accessed 29 June 2009
 57. NEA (2006) Releasing the sites of nuclear installations. A status report. Report no. NEA 6187. Nuclear Energy Agency, Issy-les-Moulineaux
 58. IAEA (2008) Long term preservation of information for decommissioning projects. Technical reports series no. 467. IAEA, Vienna, Austria. http://www-pub.iaea.org/MTCD/publications/PDF/trs467_web.pdf. Accessed 12 June 2009
 59. IAEA (2002) Record keeping for the decommissioning of nuclear facilities: guidelines and experience. Technical reports series no. 411. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/TRS411_scr.pdf. Accessed 12 June 2009
 60. Energy Information Administration (June 2006) When do commercial reactors permanently shut down? The recent record. Energy Information Administration, U.S. Department of Energy, Washington, DC. <http://www.eia.doe.gov/cneaf/nuclear/closures/closure16.html>. Accessed 14 June 2009
 61. European Commission (2009) Decommissioning in short. European Commission, Brussels. <http://www.eu-decom.be/about/decominshort/whatis.html>. Accessed 7 July 2009
 62. IAEA (2007) Decommissioning strategies for facilities using radioactive material. Safety reports series no. 50. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/Pub1281_web.pdf. Accessed 12 June 2009
 63. IAEA (2002) Safe enclosure of nuclear facilities during deferred dismantling. Safety reports series no. 26. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/Pub1142_scr.pdf. Accessed 12 June 2009
 64. USNRC (2009) Peach Bottom—Unit 1. Updated 16 Apr 2009. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/info-finder/decommissioning/power-reactor/peach-bottom-atomic-power-station-unit.html>. Accessed 12 June 2009
 65. IAEA (2008) Safety assessment for the decommissioning of facilities using radioactive material. IAEA safety standards series, Safety guide no. WS-G-5.2. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/Pub1372_web.pdf. Accessed 29 June 2009

66. Interstate Technology and Regulatory Council (Apr 2002) Determining cleanup goals at radioactively contaminated sites: case studies. Interstate Technology and Regulatory Council, Washington, DC. <http://www.itrcweb.org/Documents/RAD5.pdf>. Accessed 5 June 2009
67. IAEA (1999) Decommissioning of nuclear power plants and research reactors. IAEA safety standards series, Safety guide no. RS-G-2.1. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/P079_scr.pdf. Accessed 29 June 2009
68. USEPA (Apr 2006) Technology reference guide for radiologically contaminated surfaces. EPA-402-R-06-003. Office of Air and Radiation, U.S. Environmental Protection Agency, Washington, DC
69. IAEA (1999) State of the art technology for decontamination and dismantling of nuclear facilities. Technical reports series no. 395. IAEA, Vienna. http://www-pub.iaea.org/MTCD/publications/PDF/TRS395_scr/D395_Part1_scr.pdf. Accessed 12 June 2009
70. NEA (1999) Decontamination techniques used in decommissioning activities. Nuclear Energy Agency/Organisation for Economic Co-Operation and Development, Issy-les-Moulineaux
71. USEPA (Apr 2004) Radionuclide biological remediation resource guide. EPA-905-B-04-001. Region 5, Superfund Division, U.S. Environmental Protection Agency, Washington, DC. <http://clu-in.org/download/remed/905b04001.pdf>. Accessed 29 June 2009
72. USEPA (Apr 1999) Use of monitored natural attenuation at superfund, RCRA corrective action, and underground storage tank sites. Office of Solid Waste and Emergency Response (OSWER), Directive no. 9200.4-17P. Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency, Washington, DC. <http://www.epa.gov/swerst1/directiv/d9200417.pdf>. Accessed 29 June 2009
73. Pennsylvania Department of Environmental Protection (Oct 1998) Coal mine drainage prediction and pollution prevention in Pennsylvania. Pennsylvania Department of Environmental Protection, Harrisburg. <http://www.dep.state.pa.us/dep/deputate/minres/Districts/CMDP/main.htm>. Accessed 8 July 2009
74. IAEA (2009) Nuclear safety knowledge management. IAEA, Vienna. <http://www-ns.iaea.org/coordination/knowledge-mng.htm>. Accessed 23 July 2009
75. IAEA (2004) Planning, managing, and organizing the decommissioning of nuclear facilities. IAEA-IECDOC-1394. IAEA, Vienna
76. USNRC (Jan 2008) Decommissioning nuclear power plants. Fact sheet. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reading-rm/doc-collections/fact-sheets/decommissioning.pdf>. Accessed 14 June 2009
77. USNRC (Jan 2005) U.S. nuclear regulatory commission's decommissioning program. NUREG/BR-0325. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reading-rm/doc-collections/nuregs/brochures/br0325/br0325.pdf>. Accessed 14 June 2009

Books and Reviews

- ASTM International (2005) Standard guide for nuclear facility decommissioning plans. ASTM E1281-89(2005). ASTM International, West Conshohocken
- National Research Council (1998) A review of decontamination and decommissioning technology development programs at the department of energy. National Academies Press, Washington, DC
- National Research Council (2001) Research opportunities for deactivating and decommissioning department of energy facilities. National Academies Press, Washington, DC
- Taboas A, Moghissi A, LaGuardia T (2004) The decommissioning handbook. American Nuclear Society, LaGrange
- USDOE (Jan 2000) Decommissioning handbook. Procedures and practices for decommissioning. DOE/EM-0383. USDOE, Office of Environmental Management, Washington, DC
- Till JE, Meyer HR (eds) (Sept 1983) Radiological assessment. A textbook on environmental dose analysis. NUREG/CR-3332. U.S. Nuclear Regulatory Commission, Washington, DC
- USNRC (Aug 1988) Final generic environmental impact statement on decommissioning of nuclear facilities. NUREG-0586. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr0586/initial/index.html>
- USNRC (Nov 2002a) Generic environmental impact statement on decommissioning of nuclear facilities. Regarding the decommissioning of nuclear power reactors. NUREG-0586, suppl 1, vol 1. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr0586/s1/v1/index.html>
- USNRC (Nov 2002b) Generic environmental impact statement on decommissioning of nuclear facilities. Regarding the decommissioning of nuclear power reactors. NUREG-0586, suppl 1, vol 2. U.S. Nuclear Regulatory Commission, Washington, DC. <http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr0586/s1/v2/index.html>

Websites

- Argonne National Laboratory (USA). Decontamination and decommissioning. <http://www.dd.anl.gov/>
- Atomic Energy Canada Limited. Sustainable development: decommissioning and waste management. <http://www.aec.ca/Development/SD-WMD.htm>
- Decontamination and Decommissioning Science Consortium (USA). <http://www.orau.gov/ddsc/>
- IAEA. Decommissioning. <http://www-ns.iaea.org/tech-areas/waste-safety/decommissioning.htm>
- IAEA. Nuclear energy handbook. Technologies for the decommissioning of installations and restoration of sites. <http://www.iaea.org/inisnkm/nkm/ws/d3/r2531.html>
- NEA Working Party on Decommissioning and Dismantling (WPDD) <http://www.nea.fr/html/rwm/wpdd.html>
- Nuclear Decommissioning Authority (UK). <http://www.nda.gov.uk/>

U.S. Department of Energy (USA). Deactivation and decommissioning. <http://www.em.doe.gov/EM20Pages/DD.aspx>

U.S. Nuclear Regulatory Commission (USA). Decommissioning. <http://www.nrc.gov/about-nrc/regulatory/decommissioning.html>

World Information Service on Energy Uranium Project. <http://www.wise-uranium.org/index.html>

Nuclear Fission Power Plants

RONALD ALLEN KNIEF

Sandia National Laboratories, Albuquerque, NM, USA

Article Outline

Glossary
 Definition of the Subject
 Introduction
 Principles for Design and Operation
 Reactor Systems
 Operation, Maintenance, and Control Practices
 Nuclear Reactor Safety Features
 Nuclear-Power Regulation
 Advanced Reactors
 Future Directions
 Bibliography

Glossary

Blanket Region surrounding the fuel core of a breeder reactor that contains fertile material to increase production of new fuel.

Brayton cycle Method used to transfer fission heat energy to gas (e.g., helium or superheated carbon dioxide) for use in a gas turbine to generate electricity.

Breeder Reactor that produces new fuel from fertile material at a faster rate than it burns fuel for energy production.

Converter Reactor that produces less new fuel from fertile material than it burns for energy production.

Coolant Liquid or gaseous medium used to remove fission heat from a reactor core.

Core Region within a reactor occupied by the nuclear fuel that supports the fission chain reaction.

Critical Condition where a fission chain reaction is stable with neutron production balancing losses at a nonzero power level.

Electron volt (eV) 1 eV is the kinetic energy obtained by an electron moving across 1 V of electric potential $1 \text{ eV} = 1.602 \times 10^{-19} \text{ J}$. Common multiples are $1 \text{ keV} = 1,000 \text{ eV}$ and $1 \text{ MeV} = 10^6 \text{ eV}$. Neutron energies from less than an eV through about 10 MeV are important in nuclear fission power plants.

Fast neutrons Neutrons of high energy, particularly those produced directly by the fission reaction ($\sim 0.1\text{--}10 \text{ MeV}$).

Fertile Material, not itself fissile, capable of being converted to fissile material following absorption of a neutron.

Fissile Material capable of sustaining a fission chain reaction.

Fissionable Nuclei capable of fission by neutrons and of participation in a fission chain reaction (category includes fissile nuclides).

Fission Process in which a heavy-metal nucleus splits into two or more large fragments, releases energy, and emits neutrons and gamma radiation.

Isotopes Different nuclides of the same chemical element, e.g., ^{235}U and ^{238}U are two of the isotopes of uranium.

Moderator Material of low atomic mass included in a reactor for the purpose of reducing the kinetic energy of neutrons.

Multiplication factor Ratio of neutron production rate to neutron loss rate value is unity for a critical system.

Nuclide Atomic nucleus with a specified number of neutrons and protons, e.g., the uranium-235 [$^{235}_{92}\text{U}$] nuclide has atomic mass number 235, 92 protons (atomic number), and $235 - 92 = 143$ neutrons.

Reactivity Fractional change in neutron multiplication referenced to the critical condition value is zero for a critical system.

Reactor Combination of fissile and other materials in a geometric arrangement designed to support a neutron chain reaction.

Steam cycle Method used to convert fission heat energy to steam that drives a turbo-generator, thus, generating electricity.

Thermal neutrons Low-energy neutrons at or near thermal equilibrium with their surroundings produced by slowing down or moderating the fast neutrons produced by fission. (Equilibrium thermal energy, e.g., is 0.25-eV at 20°C).

Definition of the Subject

Nuclear fission power plants – or nuclear power reactors – are the instruments for commercial use of nuclear energy, relying on a sustained neutron chain reaction from the fission process. Nuclear fission – splitting of heavy-metal nuclei, most importantly ^{235}U and ^{239}Pu – produces an enormous amount of energy.

Fission was discovered in 1938 on the eve of World War II. The magnitude of the energy release was recognized immediately. However, it was not until after the attack on Pearl Harbor in late 1941 that the USA began serious efforts to develop this energy source for military purposes. Germany conducted research along these same lines, but ultimately was unsuccessful in development efforts.

The USA's ensuing "Manhattan Project" [1] included monumental basic research, development of nuclear reactors for research and nuclear-material production, design of nuclear explosives, and ultimately testing of a nuclear explosive device and development of the two nuclear weapons credited with ending the war. Post war, the next use of nuclear energy was for propulsion in submarines, first by the USA, then other countries. Subsequently nuclear propulsion has been extended to surface vessels as well. About 85% of the world's nuclear electricity is generated by reactors derived from designs originally developed for naval use [2].

Nuclear power for nonmilitary, peaceful uses began in earnest in 1956 with the startup of the world's first full-scale nuclear electric generating plant, at Calder Hall in England, followed in the USA a year later by the initial operation of a 60-MW unit at Shippingport, Pennsylvania [3]. By the middle of the 1960s, there was a growing optimism on the part of US utilities toward nuclear power as well as substantial activity worldwide. Electric generation using nuclear steam-supply systems was economically competitive with fossil-fuel fired plants. The passage in mid-1964 of a US law providing

for private ownership of nuclear fuel was hailed as paving the way for an independent nuclear power industry.

By 1970, the combined capacity of US nuclear plants in service, under construction, and on order was over 64 GW, more than twice the total for all other nations combined. Widespread enthusiasm for nuclear power had reached its high-water mark.

Subsequently nuclear technology suffered "growing pains." High interest rates affected the capital-cost-intensive reactors. Then in 1979 came the accident at Three Mile Island. As a result of the accident – and that at Chernobyl in 1986 [4] – over 100 orders for new reactors were cancelled and no new orders were placed in the ensuing three decades.

Despite residual public misgivings over Three Mile Island and Chernobyl, the industry has learned its lessons and established a solid safety record during the past two decades. Meanwhile efficiency and reliability of nuclear plants have climbed to record levels [5]. By 2010, 104 nuclear reactors were producing a fifth of the USA's total electric output.

"Nuclear power supplies a sixth of the world's electricity. Along with hydropower (which supplies slightly more than a sixth), it is the major source of 'carbon-free' energy today" [4]. "Nuclear power now accounts for 73% of the emission-free generation in the USA and is the only technologically mature, non-emitting source of power that is positioned to deliver large-scale CO₂ reduction in the decades ahead" [6].

Growing concern about global warming – and the associated likelihood that greenhouse gas emissions will be regulated in some fashion – has led power providers in the USA and elsewhere, and their governments [4], to recognize that nuclear reactors produce electricity without atmospheric discharge of carbon dioxide, pollutants such as nitrogen oxides, and smog-causing sulfur compounds [5]. With the world demand for energy projected to rise by about 50% by 2030 and to nearly double by 2050, building nuclear power plants appears to be increasingly viable and attractive.

The fossil-fuel alternatives have their drawbacks. Oil and natural-gas supplies are shrinking and their use can have unwanted political consequences [3]. Natural gas – attractive because it has the lowest carbon content of the fossil fuels and advanced power plants have low capital costs – has electricity costs which are

very price-sensitive and can be volatile. Although abundant, coal is difficult to recover, deliver, and burn [3, 4]. As the most carbon-intensive source of electricity, coal's long-term viability will depend on capture and sequestration of carbon dioxide – which is both costly and yet to be demonstrated and introduced on a large scale [4].

Solar, geothermal, wind, and tidal power, while attractive, cannot be expected to supply energy in the vast amounts required in the immediate future. Conservation and tight management of energy consumption are necessary, but not sufficient, parts of the equation [3].

With the future uncertain, energy realists see nuclear power as indispensable for electric generation well past the mid twenty-first century. It can fill uniquely a gap in the spectrum of electricity generation methods.

Designers are adopting novel approaches for new nuclear systems. They are evaluating systems in terms of their sustainability – meeting present needs without jeopardizing the future generations. “It is a strategy that helps to illuminate the relation between energy supplies and the needs of the environment and society. This emphasis on sustainability can lead to the development of nuclear energy-derived products besides electric power, such as hydrogen fuel for transportation. It also promotes the exploration of alternative reactor designs and nuclear fuel-recycling processes that could yield significant reductions in waste while recovering more of the energy contained in uranium” [5].

There are indications of a possible nuclear revival. More than 20,000 megawatts of nuclear capacity have come on-line globally in the early years of the twenty-first century, mostly in the Far East. Nuclear plants have demonstrated remarkable reliability and efficiency recently. The world's ample supply of uranium could fuel a much larger fleet of reactors than exists today throughout their 40- to 50-year life span [4]. The nuclear power industry has been developing and improving reactor technology for more than five decades and is starting to build the next generation of nuclear power reactors to fill orders now materializing [2]. The first firm order for new reactors in the USA was placed in 2009. Nearly 20 more considered “firmly planned” are expected to follow. There are even more prospects worldwide, most notably in pacific-basin countries [7].

Wide-scale deployment of nuclear power technology would seem to offer substantial advantages over other energy sources. However, there also are challenging questions to be addressed regarding the best way to make it fit into the future. These include achievement of economic viability, improved operating safety, efficient waste management and resource utilization, as well as weapons nonproliferation, all of which are influenced by the design of the nuclear reactor system that is chosen and the fuel cycle followed [4, 5].

Introduction

The current nuclear power reactors produce electricity using a steam cycle in much the same way as conventional fossil-fuel power plants. However, special design and operating features are required to address the unique characteristics of the fission-energy source and its associated radiation environment. A wide variety of reactor concepts have been developed to this purpose [8]. (The other major nuclear reaction for energy production, *fusion*, offers the prospect as a future energy source. See ► [Nuclear Fusion](#).)

Fission

Fission reactions occur when a neutron strikes a nucleus of specific heavy-metal isotopes of uranium and plutonium – most notably ^{235}U or ^{239}Pu – and causes splitting into two or more fission fragments, emits additional neutrons and other radiations, and releases a relatively large amount of energy. The energy release from a single fission is between five and six orders of magnitude greater than that from a single chemical reaction (e.g., combustion of carbon and oxygen). This large amount of energy per unit mass of fuel is a major advantage of fission as an energy source. The emission of extra neutrons is a necessary feature to support a sustained chain reaction with steady or accelerated energy release.

The fission reaction also has unique problems in the form of particulate and electromagnetic radiations emitted at the time of fission and from the long-lingering radioactivity (i.e., emission of radiations over time) of the fission fragments and their products. Protection against the effects of these radiations requires shielding and removal of the associated heat.

The fission process produces several hundred different fission fragments, each of which is radioactive and undergoes successive decays prior to reaching stability. Each species has a different characteristic lifetime, described in terms of its unique *half-life* – the average time required for one half of a mass of the radioactive material to decay. The resulting overall radiation levels are substantial, having an energy equivalent to about 7.5% of the total fission output at the time a reactor is shut down, and decreasing roughly inversely as the one-fifth power with time (i.e., $t^{-1/5}$). Handling this lingering decay heat load is a principal consideration in the safe operation of nuclear reactors. Containment of the radioactive material requires both near-term and long-term nuclear waste management strategies. (See ► [Health Physics](#), ► [Fission Reactor Physics](#), and ► [Radioactive Waste Management: Storage, Transport, Disposal](#).)

Fission Chain Reaction

When the chain reaction exactly balances the rates of neutron production from fission with non-fission absorption and leakage from the system boundaries, it is steady and said to be *critical*. When production exceeds losses, it is *supercritical* and power increases. When losses exceed production, the chain reaction is *subcritical* and power decreases, down to and including being shut down. All three states of criticality are necessary to nuclear power reactor operation. These states are quantified with an *effective multiplication factor* k_{eff} or k , defined as

$$k_{\text{eff}} = k = \text{production} / [\text{absorption} + \text{leakage}]$$

or by the change in multiplication, *reactivity* ρ , defined as

$$\rho = (k - 1) / k.$$

Thus, $k = 1$ and $\rho = 0$ constitute the critical condition; $k > 1$ and $\rho > 0$ supercritical; and $k < 1$ and $\rho < 0$ subcritical.

The mathematical equations associated with the neutron balance are quite complex since the reaction rates are a strong function of both material (composition, location, and irradiation history) and neutron population (position, direction, and energy) characteristics. Neutron energy is a particularly important

parameter. Neutrons emitted in fission are in the 0.1–10 MeV range with an average energy of about 2 MeV. The probability for fission increases dramatically if the neutron energy falls below 1 eV.

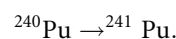
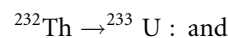
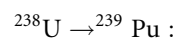
For this reason, a *moderator* material of low mass may be added to enhance the slowing down. (The closer the mass of the moderator nuclei to that of the neutron, the larger the potential energy loss in a single collision; thus, hydrogen with essentially the same mass is best in this regard, although it may not be selected for other reasons.) (See ► [Nuclear Reactor Materials and Fuels](#).)

Fissionable Materials

The heavy metals uranium and plutonium each have isotopes that can be caused to fission by neutrons, i.e., that are *fissionable*. Fission occurs under different circumstances among the isotopes.

Fissile nuclides can fission with neutrons of energy, all the way down to essentially zero. The principal fissile materials are those containing significant quantities of ^{233}U , ^{235}U , ^{239}Pu , and/or ^{241}Pu . A fission reactor core contains fissile as well as non-fissile and other materials.

Other isotopes of U, Pu, and Th can be fissioned by neutrons, but only by neutrons with kinetic energy above a minimum threshold. Particularly important examples of such *fissionable* nuclides are ^{238}U , ^{232}Th , and ^{240}Pu . The latter nuclides also have the unique characteristic that upon absorbing a neutron in a non-fission reaction, they will convert to fissile nuclides either directly or via a multi-step process. These *fertile* nuclides produce the new fissile as follows:



(See ► [Fission Reactor Physics](#))

Reactor Classifications

Nuclear power reactors are designed to achieve a self-sustained chain reaction with a combination of fissile, fertile, and structural materials. The current generation of reactors uses the heat produced by fissions in the core to produce steam and then electricity, employing

a variety of fuel forms, coolants, moderators, and other materials.

Common characteristics useful for classifying reactors are:

1. *Coolant* – principal heat-removal medium.
2. *Steam cycle* – number and configuration of separate coolant “loops” for producing steam to turn a turbine generator and produce electricity.
3. *Moderator* – material (if any) used to “slow down” the neutrons produced by fission.
4. *Neutron energy* – general energy range (typically fast [fission-energy] or thermal) for the neutrons that cause most of the fissions.
5. *Fuel production* – system is referred to as a *breeder* if it produces (i.e., changes from fertile to fissile) more fuel than it consumes; it is said to be a *converter* otherwise.

The first two features relate to the current practice of converting fission energy first to heat and then to electric energy by employing a steam cycle. Coolants include water, heavy water, gases (e.g., CO₂ or helium), and liquid metal (e.g., sodium). The steam cycles may employ from one to three separate loops, including one for primary coolant circulation and one (not necessarily separate from the first) for steam generation. It may be noted that next-generation reactors could bypass steam in favor of direct use of a coolant for electricity, process heat, or hydrogen production.

As mentioned earlier, neutrons emitted from fission are of high energy. However, very low energy neutrons have a higher likelihood of causing fission reactions. Thus, many systems employ a moderator to “slow down” these neutrons. The best moderators are of low mass, allowing maximum energy transfer through neutron collisions (e.g., the limiting case of potentially total energy transfer between a moving cue ball and a stationary billiard ball of equal mass). Typical materials used for this purpose are hydrogen, deuterium (heavy hydrogen), and carbon/graphite. The moderator and coolant may be the same (e.g., water or heavy water) or may be separate (e.g., water, heavy water, or a gas coolant with heavy water or solid graphite moderator). Neutrons with low enough energies to be roughly in thermal equilibrium (<1 eV) with the surrounding materials are said to be *thermal neutrons*. Neutrons at or near fission energies (averaging

~2 MeV) are *fast neutrons*. Fast reactors avoid the use of moderators, such as with a metal coolant like sodium, instead of one of the moderating materials identified above.

All nuclear power reactors contain fertile ²³⁸U or ²³²Th and, thus, produce some amount of new fissile fuel. The ubiquitous light-water reactors (LWR) have slightly enriched uranium fuel which is up to 5 wt% ²³⁵U with the remaining 95 wt% ²³⁸U. Thus, they produce plutonium during routine operation. Producing about 60% as much new fuel as used for operation, the LWR units are considered to be *converter* reactors. By contrast, a reactor producing more new fuel than it uses is called a *breeder*.

The world’s six principal types in the current generation of nuclear power reactors are:

1. Boiling-water reactor (BWR)
2. Pressurized-water reactor (PWR)
3. Heavy-water-moderated reactor (HWR)
4. Gas-cooled reactor (GCR)
5. Light-water-cooled graphite-moderated reactor (LWGR)
6. Fast breeder reactor (FBR)

The section labeled “General” in Table 1 classifies one or more versions of each of these six reactor types in terms of the five categories identified at the beginning of this section [8]. For example, the GCR variation known as the high-temperature gas-cooled reactor (HTGR) is a double-loop, helium-cooled, graphite-moderated, thermal, converter reactor.

Principles for Design and Operation

Nuclear power reactors are complex systems whose design represents a balance among conflicting requirements. Principal among these requirements are nuclear design, materials, thermal design, economics, and control and safety [8, 9].

The nuclear (physics) design seeks to match fissile and fertile constituents with appropriate coolants and moderators (if any) to optimize specific (per unit mass of fuel) fission-energy output and production of new fuel or meet alternative design goals. Materials concern focus on chemical compatibility of components, thermal and radiation stability, and overall mechanical strength. One especially important requirement for

Nuclear Fission Power Plants. Table 1 Characteristics for six representative nuclear steam-supply systems

	BWR	PWR	PHWR	PTGR	HTGR	LMFBR
	Manufacturer					
	General Electric	Westinghouse	Atomic Energy of Canada, Ltd.	(Former) Soviet Union	General Atomic	Novatome
	System (Station)					
	BWR/6	(Sequoyah/ SNUPPS)	CANDU 600	RBMK-1000	(Fulton)	(Superphénix)
General						
Steam cycle						
Loops	1	2	2	1	2	3
Primary coolant	H ₂ O	H ₂ O	D ₂ O	H ₂ O	He	Liq. Na
Secondary coolant	–	H ₂ O	H ₂ O	–	H ₂ O	Liq. Na/H ₂ O
Moderator	H ₂ O	H ₂ O	D ₂ O	Graphite	Graphite	–
Neutron energy	Thermal	Thermal	Thermal	Thermal	Thermal	Fast
Fuel production	Converter	Converter	Converter	Converter	Converter	Breeder
Energy conversion						
Gross thermal power, MWt	3,579	3,411	2,180	3,200	3000	3,000
Net electric power, MWe	1,178	1,150	638	950	1,160	1,200
Efficiency, %	32.9	33.7	29.3	31.2	38.7	40
Heat transport						
Primary loops and pumps	2	4	2	2/6 + 2	6	4
Intermediate loops	–	–	–	–	–	8
Steam generators	–	4	4	–	6	8
Steam-generator type	–	U-tube	U-tube	–	Helical coil	Helical coil
No. turbine generators	1	1	1	2	1	2
Fuel						
Particles	Short, cyl. pellets	Short, cyl. pellets	Short, cyl. pellets	Short, cyl. pellets	Coated microspheres	Short, cyl. pellets
Chemical form	UO ₂	UO ₂	UO ₂	UO ₂	UC/ThC	Mixed UO ₂ /PuO ₂
Fissile (% by wt.)	2–4% ²³⁵ U	2–4% ²³⁵ U	Natural uranium	Natural uranium	93% ²³⁵ U	10–20% Pu (core)
Fertile	²³⁸ U	²³⁸ U	²³⁸ U	²³⁸ U	Thorium	²³⁸ U (core + blanket)

Nuclear Fission Power Plants. Table 1 (Continued)

	BWR	PWR	PHWR	PTGR	HTGR	LMFBR
	Manufacturer					
	General Electric	Westinghouse	Atomic Energy of Canada, Ltd.	(Former) Soviet Union	General Atomic	Novatome
	System (Station)					
	BWR/6	(Sequoyah/SNUPPS)	CANDU 600	RBMK-1000	(Fulton)	(Superphénix)
Pins	Pellet stacks in Zr-alloy tubes	Pellet stacks in Zr-alloy tubes	Pellet stacks in Zr-alloy tubes	Pellet stacks in Zr-alloy tubes	Microspheres in graphite stick	Pellet stacks in SS tubes
Assembly	8 × 8 sq. pin array	17 × 17 sq. pin array	37-pin concentric circles	18-pin concentric circles	Hex. graphite block with fuel sticks	271-pin hex. array
Core						
Axis	Vertical	Vertical	Horizontal	Vertical	Vertical	Vertical
Assys on axis	1	1	12	2	8	1
Assys radially	748	193	380	1,661	493	364 (core), 233 (blanket)
Performance						
Equil. burnup, MWD/T	27,500	27,500	7,500	18,500	95,000	100,000
Refueling sequence	1/4 per year	1/3 per year	Continuous, on-line	On-line	1/4 per year	Variable
Thermal hydraulics						
Primary system						
Pressure, MPa	7.17	15.5	10.0	7.2	4.90	~0.1
Inlet temp., °C	278	292	267	270	318	395
Avg. out. temp., °C	288	325	310	284	741	545
Core flow, Mg/s	13.1	18.0	7.6	10.4	1.42	16.4
Volume, L	–	3.36×10^5	1.20×10^5	4.0	(9,550 kg)	(3,200 Mg) Na/H ₂ O
Secondary system						
Pressure, MPa	–	6.89	4.7	54	17.2	~0.1/17.7
Inlet temp., °C	–	227	187	29	188	345/235
Outlet temp., °C	–	285	260		513	525/487
Power density						
Core avg., kW/L	54.1	105	12	4.0	8.4	280
Fuel avg., kW/L	54.1	105	60	54	44	280

Nuclear Fission Power Plants. Table 1 (Continued)

	BWR	PWR	PHWR	PTGR	HTGR	LMFBR
	Manufacturer					
	General Electric	Westinghouse	Atomic Energy of Canada, Ltd.	(Former) Soviet Union	General Atomic	Novatome
	System (Station)					
	BWR/6	(Sequoyah/ SNUPPS)	CANDU 600	RBMK-1000	(Fulton)	(Superphénix)
Linear heat rate						
Core avg., kW/m	19.0	17.8	25.7	29	7.87	29
Core max., kW/m	44.0	42.7	44.1		23.0	45
Design peaking factors						
Radial	1.4		1.21			
(Total)		(2.5)			(2.9)	(1.55)
Axial	1.6		1.41			
Moderator						
Volume, L	(Same as primary coolant)	(Same as primary coolant)	2.17×10^6	(Graphite)	(Graphite in fuel blocks)	–
Inlet temp., °C			43			–
Outlet temp., °C			71			–
Reactivity control						
Control rods						
Geometry	Cruciform	Rod clusters	Rods	Rods	Rod pairs	Hex. pin bundle
Absorber material	B ₄ C	Ag In Cd	Various	B ₄ C	B ₄ C/graphite	B ₄ C
Burnable poison	Gd. in fuel pellets	Borosilicate glass	–		B ₄ C/graphite	–
Other systems	Voids in coolant	Soluble boron	H ₂ O, various		Reserve shutdown	Three bundle secondary
Reactor vessel						
Inside dimension, m	6.05D × 21.6 H	4.83D × 13.4 H	7.6D × 4 L	0.08ID × 8 H tubes	11.3D × 14.4 H	21D × 19.5 H
Wall thick., mm	152	224	28.6	4.0	(4.72 m min.)	25
Material	SS-clad carbon steel	SS-clad carbon steel	Stainless steel	Zr-Nb alloy	Prestressed concrete	Stainless steel
Other features			Pressure tubes	Pressure tubes	Steel liners	Pool-type

Data summarized from [8]. Appendix IV. (Appendix also includes parameters for PWR designs from US manufacturers B&W and CE, France's Framatome, and the former Soviet Union's unique VVER)

reactors that refuel only while shutdown – which is the majority – is that the fuel maintains its general structural integrity throughout up to 4 years or more of in-core residence; unlike other energy production cycles, the fuel is not literally “burned up” with wastes discarded along the way.

Thermal design is concerned with heat removal for effective energy production and maintaining the integrity of the fuel, cladding, and other materials for operational and safety purposes. Reactor thermal operating limits are determined by the most extreme postulated local conditions – to avoid damage anywhere in the reactor core. Temperature limits are established to prevent fuel melting and/or clad damage during power excursions and loss of cooling. Other limits address avoiding coolant-clad temperature mismatch that could lead to local loss of fuel integrity. Both types of limits are established in terms of peak (local)-to-average core conditions. Such factors vary with power level, time in core life, and other parameters.

The primary economic consideration is to minimize overall costs, i.e., initial capital outlay, operating and maintenance costs, and fuel charges. Increasing electric-generation reliability and thermal conversion efficiency are particularly important throughout lifetime.

Power reactor controls must be capable of maintaining the critical condition, increasing and decreasing power, and adjusting to long-term changes such as the conflicting effects from breeding new fuel, depleting existing fuel and burnable neutron poisons, and building in radioactive fission- and transuranic-product neutron poisons. The desired neutron balance is maintained predominately by adjusting neutron absorption using materials designed to remove neutrons from, or “poison,” the chain reaction. Methods may include a combination of solid moveable control rods, soluble poisons in the coolant or moderator (PWR), and fixed burnable poisons designed to deplete or be “burned out” by the continuing neutron population. Some reactor designs also change neutron production by on-line fuel exchange. Fission products whose poisoning effects are sensitive to operating conditions also must be subject to control.

Routine control strives to make the power density as uniform as possible throughout the core, while allowing for power changes. In most designs, control-rod movement is used with groups selected for symmetry to

maintain uniform, i.e., avoid highly asymmetric, power distribution. Measures are instituted to restrict the speed of movement and reactivity effect of individual rods or groups of rods to prevent excessively rapid power increase. Similarly, the design intends to minimize the likelihood of inadvertent control-rod withdrawal which could lead to unplanned supercritical conditions.

Safety concerns are addressed through a protective system whereby the control rods may be inserted into the core – dropped down or pushed up – into the core quickly; i.e., they “scram” or “trip” the reactor through gravity drop or gas pressure, respectively, when certain predetermined parameter limits (e.g., on pressure, temperature, flow, or power levels) are exceeded.

Reactor safety is enhanced with inherent negative feedback mechanisms, where a power increase tends to be self-limiting. Fuels are designed to have such feedback, with their reactivity decreasing with escalating temperature. Expansion of the coolant and/or moderator (such as with increased temperature) provides negative reactivity feedback in light-water reactors. Reactors with positive reactivity feedback in certain operating regimes may be difficult to control. The essentially instantaneous feedback contributions, along with some long-term ones (such as from time-dependent changes in concentrations of xenon and samarium fission-product poisons), affect the stability of the chain reaction, may limit the ability to change the power level, and are a major consideration in overall reactor safety design.

A fundamental safety feature for all power reactors is multiple-barrier containment of fission products. As may be observed for each reactor type described in the remainder of this article, these barriers include the fuel particles, surrounding cladding, the coolant system boundary, and a containment structure.

An important example of tradeoffs among the design goals is seen in thermal-reactor fuel assemblies (e.g., subsequent Figs. 3 and 8) whose pin arrangement determines the characteristics of the chain reaction, economics, and heat removal [9]. The chain reaction is enhanced by optimum spacing of the fuel in “lumps” with moderator interspersed so that neutrons from fission will undergo a number of scattering collisions for slowing down prior to reentering the fuel; too little and too much spacing can both be detrimental.

The extent of slowing down also determines the amount of conversion of fertile to fissile material and the overall energy production possible from a given amount of fuel. Spacing and coolant flow rate establish heat-removal characteristics (including temperature feedback effects on the fuel's reactivity). Final dimensions generally represent a best-estimate balance among these and other competing concerns.

Reactor Systems

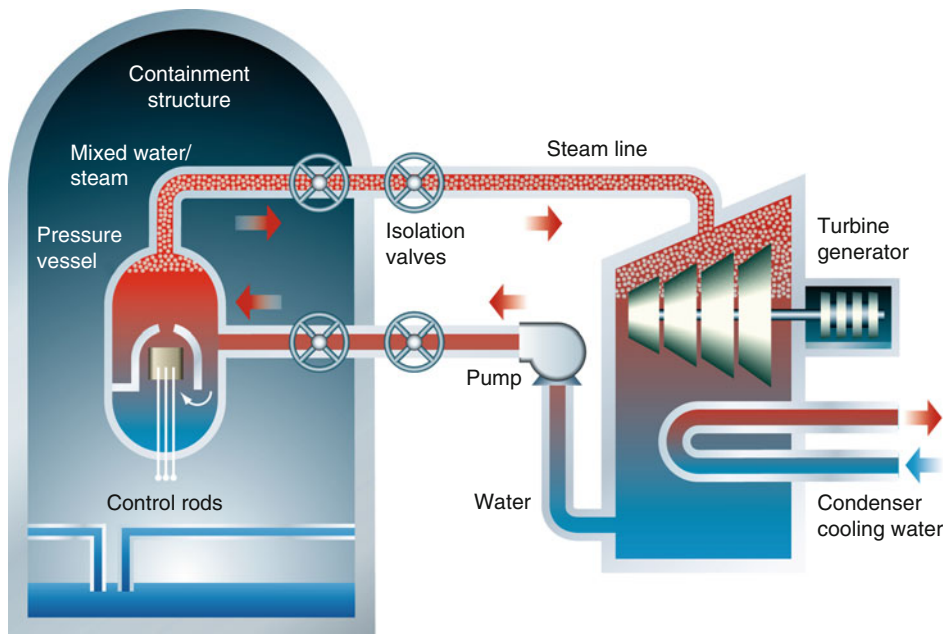
The major nuclear power reactor types were identified at the end of the section “[Introduction](#)” noting that characterization of each is provided in the “General” section of [Table 1](#). More detailed descriptions of each of these systems as implemented in the present generation of reactors follow. (As explained in detail in the section “[Advanced Reactors](#),” several generations of reactors are commonly distinguished. Of interest here, Generation I (GEN-I) reactors were developed in the 1950s–1960s, while Generation II (GEN-II) reactors are typified by the present US fleet and most in operation elsewhere.) Parameters for selected specific present reactor designs are detailed in the remainder of [Table 1](#).

The main focus of these nuclear power plant descriptions is on the steam cycle, fuel assemblies, reactivity control, and the protective system [8]. General safety-related functions are summarized separately in the section “[Nuclear Reactor Safety Features](#).”

Light-Water Reactors

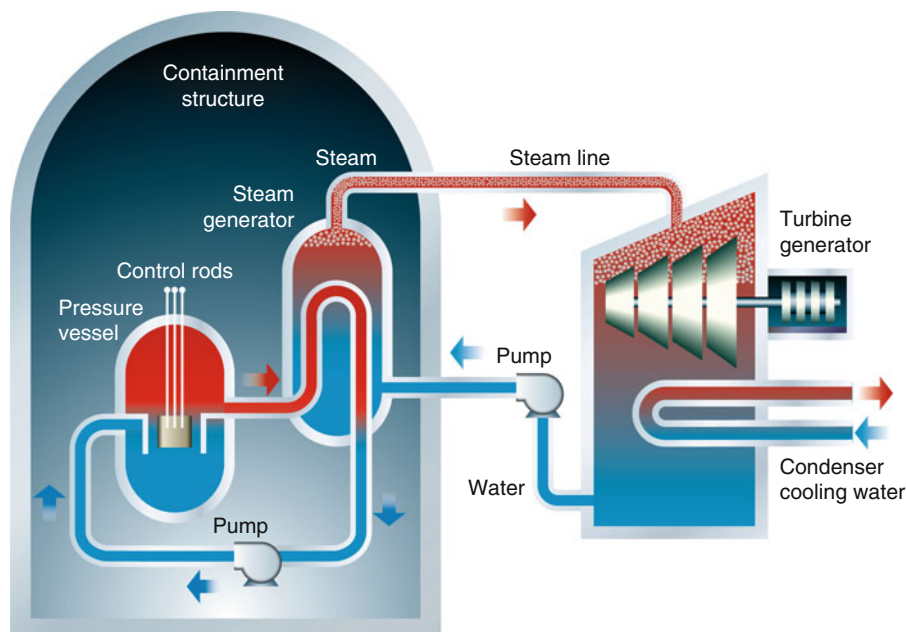
Two light-water reactor (LWR) systems – boiling-water reactor (BWR) and pressurized-water reactor (PWR) – employ ordinary (“light”) water as coolant and moderator. The former design produces steam through a direct cycle ([Fig. 1](#)), while the latter uses an intermediate steam-generator heat exchanger to maintain an all-liquid primary loop and produce steam in a separate secondary loop ([Fig. 2](#)).

The nature of the water coolant/moderator results in similarities between the two LWR designs. The fuel is uranium dioxide pellets enriched to 2–5 wt% ^{235}U ; initial core loadings typically span this range with reload cores distinctly at the upper end of the enrichment range. The pellets are clad in sealed zirconium-alloy tubes. Fuel assemblies consist of rectangular arrays of fuel pins with regular spacing (see also ► [Modern Nuclear Fuel Cycles](#)).



Nuclear Fission Power Plants. Figure 1

Steam cycle for boiling-water reactor (BWR) (Courtesy of Atomic Industrial Forum/Nuclear Energy Institute)



Nuclear Fission Power Plants. Figure 2

Steam cycle for pressurized-water reactor (PWR) (Courtesy of Atomic Industrial Forum/Nuclear Energy Institute)

Since the LWR designs rely on liquid water for moderating neutrons, maximum operating temperatures must remain well below the 706°F (374°C) critical temperature at which pressure increases dramatically and liquid cannot exist regardless of pressure. They, thus, cannot operate at the “modern steam conditions” – nominally 1,000°F (540°C) – typical of fossil-fueled plants. The LWR’s saturated steam–water mixture requires a special, more costly “wet steam” turbine.

Boiling-Water Reactors The direct-cycle boiling-water reactor (BWR) was developed in the USA by General Electric and was followed by designs in Western Europe and Japan. It may be noted that the current designs for BWR and other reactors are increasingly characterized by multinational joint ventures. See also the section “[Advanced Reactors](#).” For details on historical, current, and proposed units of this and the other designs, see Refs. [7, 8, 10, 11].

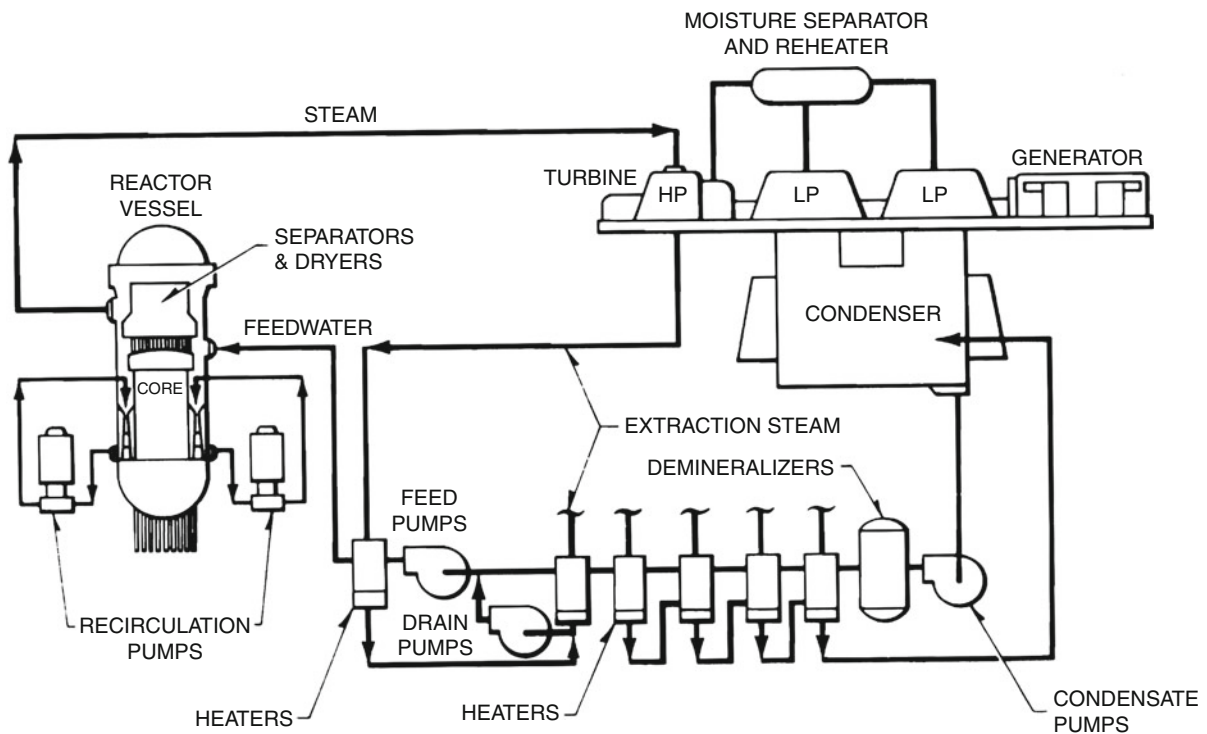
Employing the cycle shown in concept by [Fig. 1](#) and in more detail in [Fig. 3](#), feedwater enters the steel reactor vessel, is heated by the fission chain reaction occurring in the fuel pins, and leaves the vessel as

steam. The high- and low-pressure turbine stages are employed in concert with the multiple heaters and condensers to enhance energy-conversion efficiency. The more recent BWR designs use jet pumps to recirculate a fraction of the feedwater flow for better control.

Fuel assemblies for the BWR appear as shown in [Fig. 4](#). The 7×7 to 9×9 square arrays of fuel pins is surrounded by a metal fuel channel, which prevents the water–steam mixture from moving between assemblies (and potentially resulting in inadequate cooling (steam–water mixture) in some assemblies). Fuel assemblies may contain pins of several different enrichments ([Fig. 5](#)). The reactor fuel core consists of up to 800 fuel assemblies.

Reactivity control for routine operation is implemented through a combination of control rods and coolant flow adjustment. The bottom-mounted control rods (indicated below the reactor vessel in [Fig. 3](#)) are made of long boron carbide–filled pins in a cruciform (“cross”) shape that fits between four fuel assemblies as shown in [Fig. 5](#).

Flow adjustment can provide another effective control method, since the water density changes with



Nuclear Fission Power Plants. Figure 3

Representative steam cycle schematic diagram for a boiling-water reactor (BWR) (Courtesy of General Electric Company)

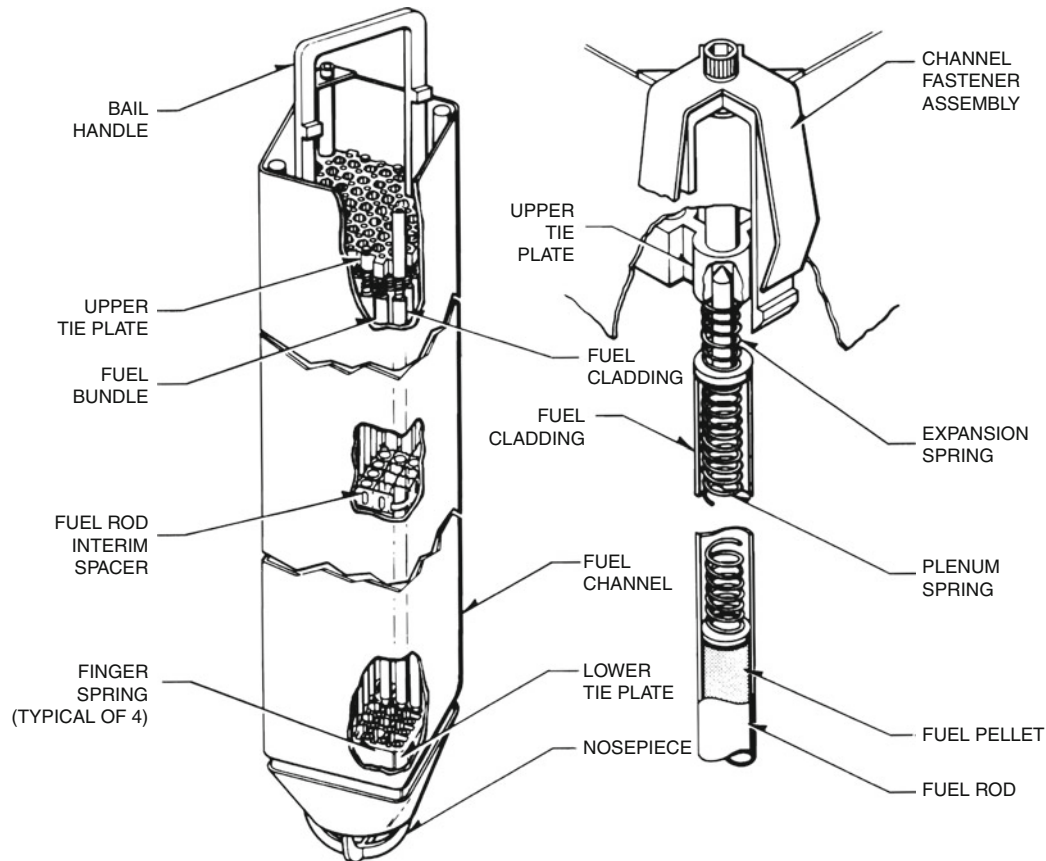
temperature. At low temperature, the dense water is very effective at moderating neutrons, and thereby encourages fission. With increased temperature, the density decreases (or, equivalently, void content increases as steam is being produced), causing a reduction in moderation and fission rate. Thus, if flow rate is increased, energy removal can be increased without a net change in coolant temperature with a resulting increase in power generation. In practice, power-level changes of up to 40% may be accomplished by flow control. (BWR operational practices are described further in the section “[Boiling-Water Reactor](#).”)

Longer-term reactivity control is accomplished using burnable poisons (e.g., “curtains” containing boron placed between fuel assemblies or gadolinium poisons fabricated into the fuel itself) and gradually withdrawing inserted control rods over core lifetime. Reactor shutdown – scram or trip – is accomplished by using gas pressure to insert all of the bottom-mounted control rods into the core.

[Table 1](#) contains parameters for a large BWR and other reactor types. Advanced BWRs are addressed in the section “[Advanced Reactors](#).”

Pressurized-Water Reactor The two-loop pressurized-water reactor (PWR) was developed in the USA by Westinghouse Electric Corporation. Other somewhat similar designs have been developed by a pair of US organizations, as well as by others in Western Europe, Japan, and South Korea. A particular unique version has come from the Soviet Union (now by Russia). Water in the primary loop ([Fig. 2](#)) is maintained as liquid by using high pressure. The water enters the reactor vessel ([Fig. 6](#)) at the inlet nozzle, flows downward along the inner vessel wall, is distributed at the lower vessel plate, flows up through the fuel assemblies removing heat energy, and exits at the outlet nozzle – still as a liquid.

Energy from the primary loop is extracted and converted to steam by two to four U-tube ([Fig. 7](#)), once-through, or horizontal (Russian design) steam



Nuclear Fission Power Plants. Figure 4

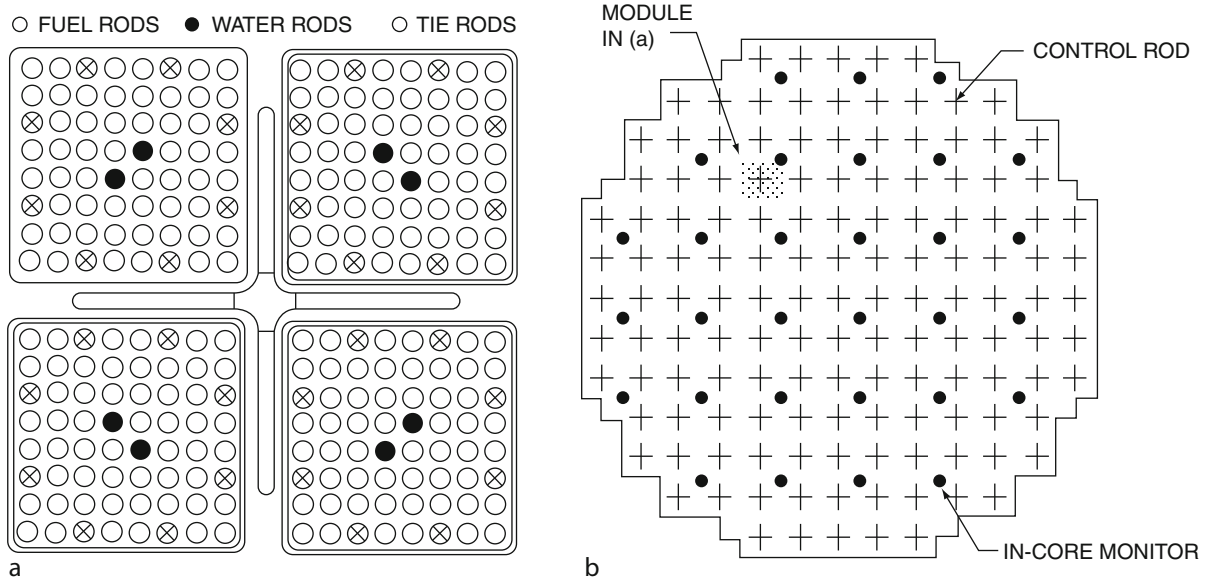
Representative fuel assembly for boiling-water reactor (Courtesy of General Electric Company)

generators. Multiple turbine stages, heaters, and a condenser are employed as for the BWR (Fig. 3). A *pressurizer* with a steam–water interface is used to maintain the sensitive pressure/temperature balance in the primary system by using heaters to boil water, increase the amount steam, and thus raise the pressure; conversely spraying “cooler” (“less hot”) water condenses steam and reduces the pressure.

Fuel assemblies for the PWR are of 14×14 to 17×17 square fuel pin arrays (Fig. 8) or a hexagonal array of up to 331 pins (Russian VVER). They are not enclosed in a fuel channel (in part because the single-phase primary fluid is better behaved than the BWR’s boiling coolant). These assemblies also have unoccupied pin locations, which can accommodate control rods, burnable poisons, or instruments. The large PWR reactor cores consist of 150–200 or more fuel assemblies.

Like the BWR, the PWR uses control and burnable-poison rods for reactivity control. Uniquely, it also uses the soluble neutron poison, boric acid, for long-term control. The boric acid concentration is adjusted to match general changes from fuel burnup, conversion of fertile material into fissile, and depletion of burnable poisons.

Insertion of control-rod assemblies from the top of the core (by contrast with the bottom-mounted BWR control rods) provides short-term, routine reactivity control. These assemblies consist of 5–24 “fingers,” made of boron carbide or of a silver–indium–cadmium mixture, which move in channels within the fuel assemblies (e.g., as shown in Fig. 8). A small group of rods in a symmetric pattern is inserted partially into the fuel (typically >25%) and then moved as needed to compensate for routine



Nuclear Fission Power Plants. Figure 5

Core fuel module (a) and control-rod pattern (b) for representative boiling-water reactor (Courtesy of General Electric Company)

power fluctuations. Significant power changes up to an including startup and shutdown invoke other, larger groups of control rods. (PWR operational practices are described further in the section “[Pressurized-Water Reactor](#).”)

Scram/trip is accomplished in the PWR by dropping the top-mounted control rods (Figs. 6 and 8) into the core under the influence of gravity. The rods are mounted to their drives by electromagnets so that interruption of the current (from power failure or indication that a specific parameter is outside of a predetermined range) causes the rods to fall.

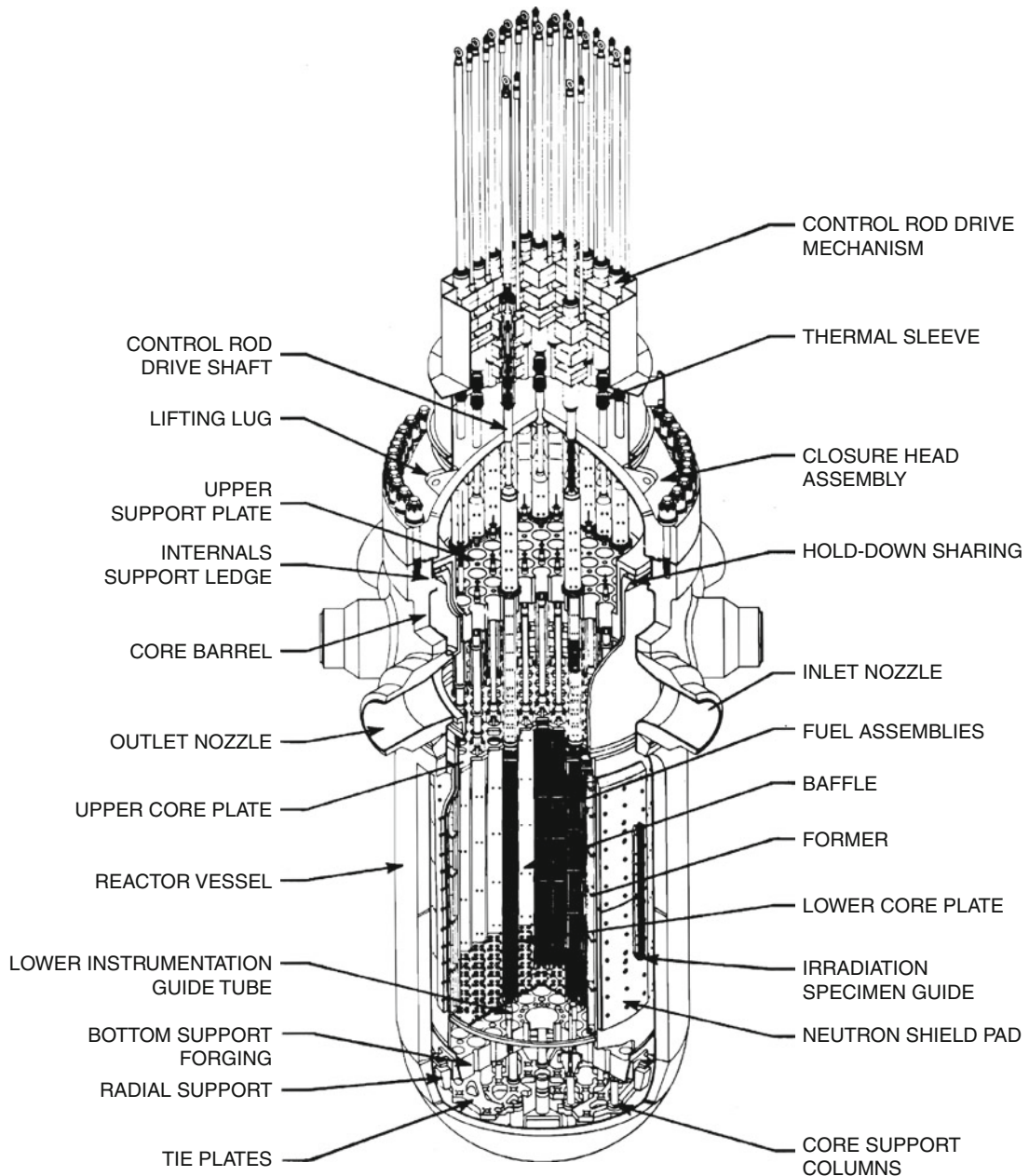
Figure 9 is a cutaway drawing (wallchart) of a representative PWR unit. Other than for the reactor “nuclear island” – also known as the nuclear steam-supply system (NSSS) – the drawing shows features common to the other reactor units, all of which have turbine generator(s), and containment, auxiliary, and support buildings. (Reference [12] contains a Web link to this wallchart, as well as to charts for over 100 other reactors.)

Table 1 contains parameters for a representative PWR and other reactor types. Advanced PWRs are addressed in the section “[Advanced Reactors](#).”

Heavy-Water Reactors

Ordinary hydrogen in the form of water is the most effective material for reducing neutron energy, but it also absorbs some of the neutrons that could otherwise participate in the chain reaction. Thus, deuterium as heavy water, which requires more collisions for a given energy change but exhibits much less absorption, is a better moderator in some ways, e.g., allowing use of natural- or other low-enrichment uranium fuel. Deuterium, existing in nature in a ratio of 1:400 with ordinary hydrogen, requires isotopic enrichment prior to use (as does ^{235}U in uranium for many of the reactor applications).

Heavy-water reactors (HWR) have been developed in Canada – their signature CANDU (for Canadian deuterium-uranium) – the UK, West Germany, India, Japan, and South Korea. Pressure-vessel designs (with similar hardware configurations to the PWR) employ the same heavy water volume as coolant and moderator. Pressure-tube designs use heavy water in a separate moderating volume generally with pressurized heavy-water (PHW) coolant, sometimes boiling light-water coolant, and potentially with an organic liquid.



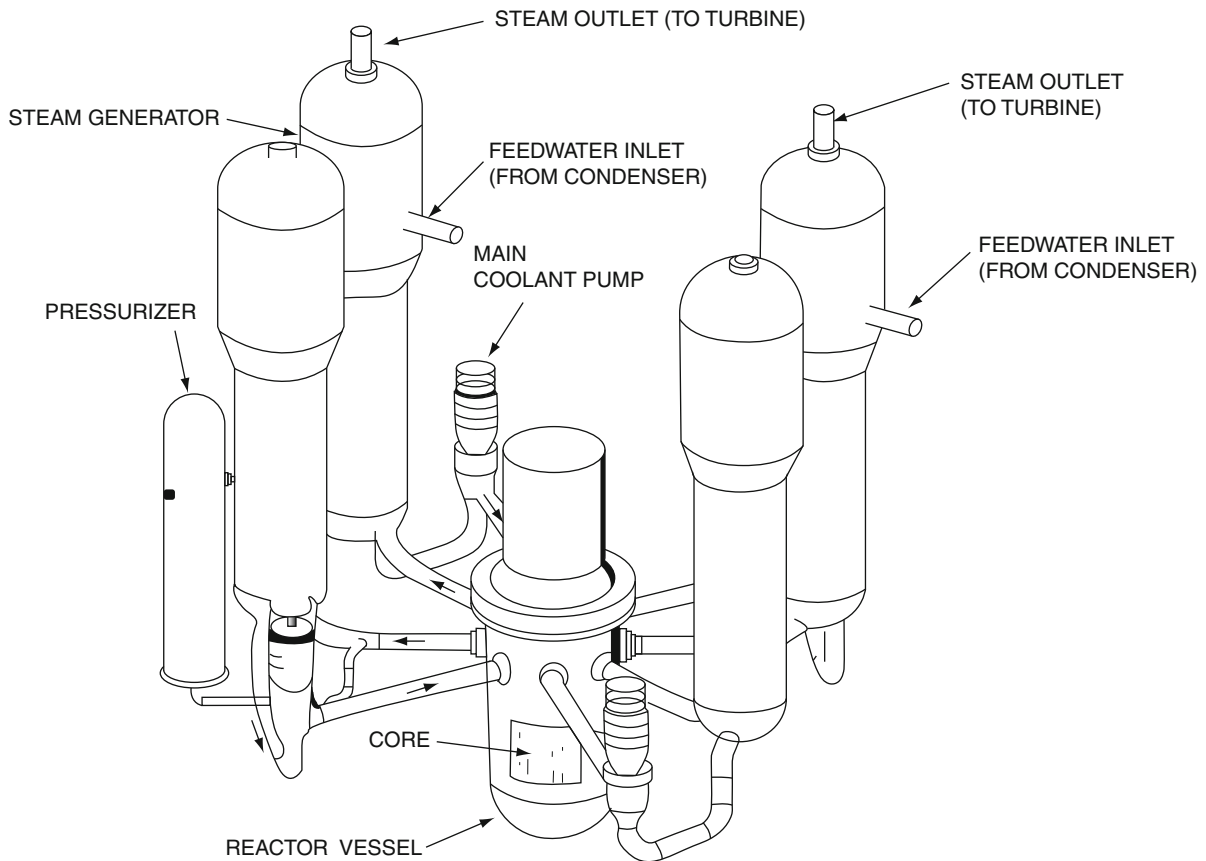
Nuclear Fission Power Plants. Figure 6

Reactor pressure vessel for a representative pressurized-water reactor (Courtesy of Westinghouse Electric Company)

The example of a CANDU-PHW is considered further below.

The CANDU-PHW steam cycle has two loops (Fig. 10), like the PWR, with the primary pressurized heavy-water loop transferring heat energy to a loop of ordinary water for steam production. A major

difference, however, is that the primary fluid is distributed among several hundred pressure tubes which pass through a large calandria vessel (Fig. 11) containing the separate heavy-water moderator. The coolant is actually collected in two separate loops, with the fluid in adjacent tubes flowing in opposite directions.



Nuclear Fission Power Plants. Figure 7

Four U-tube steam-generator primary loop configuration for a pressurized-water reactor (Courtesy of Westinghouse Electric Company)

The fuel assemblies consist of natural (i.e., 0.711 wt% ^{235}U) uranium dioxide fuel pellets in zirconium clad, similar to LWR fuel. However, short, cylindrical bundles of fuel pins (Fig. 12) allow a unique on-line fueling scheme whereby a machine attaches to each end of a single coolant tube and inserts one fuel assembly while in effect pushing out another.

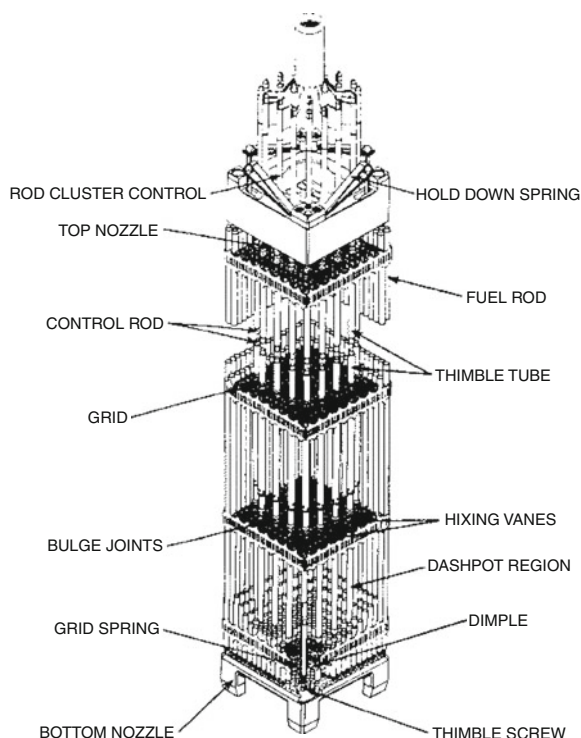
A major portion of the reactivity control is accomplished by on-line fueling, which is required to compensate for the low reactivity inherent in the natural uranium fuel. Routine operating adjustments and power shaping are accomplished with neutron-poison control rods, or by introduction of ordinary water (which absorbs more neutrons than heavy water) into special chambers. Other control rods are available for reactor scram/trip. The separation of the coolant and moderator volumes also provides the possibility for

moderator “dumping” as an emergency shutdown method.

Table 1 contains parameters for a large CANDU-PHW and other reactor types. Advanced HWRs are addressed in the section “Advanced Reactors.”

Gas-Cooled Reactors

The world’s first research reactor used natural uranium fuel, graphite moderator, and natural-circulation air cooling. Subsequent systems have also used graphite moderator with natural or enriched uranium and with carbon dioxide or helium coolant. Various commercial gas-cooled reactors (GCR) have operated in France, the UK, the USA, and Germany. Two versions of the helium-cooled high-temperature gas-cooled reactor (HTGR) developed by the USA and Germany are



Nuclear Fission Power Plants. Figure 8
Representative fuel assembly for a pressurized-water reactor (Courtesy of Westinghouse Electric Corporation)

described below. (Both are prototypical designs at $\sim 1,000$ MWe described for the purpose of comparison with other current large commercial reactors in this section. Evolutionary units of this heritage are addressed in the section “[Advanced Reactors](#).”)

The HTGR steam cycle ([Fig. 13](#)) employs a primary loop of helium, heat exchangers, and pumps contained within a prestressed concrete reactor vessel (PCRV) – two versions of which are shown in [Fig. 14](#) – and a steam loop. Since the coolant is a single-phase gas, no pressurizer is required (in contrast to the two-loop water-cooled designs). The nature of the coolant also provides the prospect for direct conversion through a gas turbine, e.g., employing a Brayton thermodynamic cycle.

Fuel for the HTGR consists of small microspheres of uranium or thorium carbide (UC/ThC) with coatings of graphite and/or silicon carbide ([Fig. 15a](#)). The uranium microspheres, enriched to 20–93 wt%, may be mixed with separate thorium microspheres to an effective fissile enrichment of about 5 wt%.

In the US “prismatic” – hexagonal-block – HTGR system, the microsphere mixture is formed into roughly finger-sized sticks with a carbon-resin binder. The sticks are then loaded into large hexagonal graphite blocks with interspersed coolant holes ([Fig. 15b](#)). The blocks are stacked several high and in a roughly cylindrical arrangement to form the reactor core ([Fig. 14a](#)). Another version of the HTGR – the thorium high-temperature reactor (THTR) – forms the microspheres into a spherical shape and coats them with hard, thick graphite layers into a 6-cm-diameter “pebble” ([Fig. 15c](#)). The reactor core is then formed by loading these fuel units into a hopper in a PCRV ([Fig. 14b](#)), from which fueling and defueling can be accomplished on-line. This THTR design feature leads to the alternative designation of the reactor as “pebble bed.”

Reactivity control in the prismatic design depends on boron control rods for routine and shutdown functions. Burnable poisons may be used for long-term reactivity control. A reserve shutdown system consisting of small boron-carbide balls backs up the primary systems.

The THTR requires minimal excess reactivity due to its ability to change fuel on-line. Control rods inserted into the pool of pebbles provide the means for routine operational adjustments.

[Table 1](#) contains parameters for a large prismatic HTGR and other reactor types. Advanced GCRs – prismatic and pebble bed – are addressed in the section “[Advanced Reactors](#).”

Light-Water Graphite Reactors

Light-water-cooled graphite-moderated reactors (LWGR)/pressure-tube graphite reactors (PTGR) were among the first systems used for purposes of research, plutonium production, and electric power generation. The Soviet Union credits a small unit of this type with generating the first commercial electricity. Current commercial use is limited to the Soviet RBMK pressure-tube graphite reactors (PTGR). The Chernobyl reactor – the site of the catastrophic 1986 accident – was of this type.

The RBMK uses a direct steam cycle ([Fig. 16](#)) with boiling-water coolant like the BWR ([Fig. 1](#)). However, its complex pressure-tube design with separate coolant



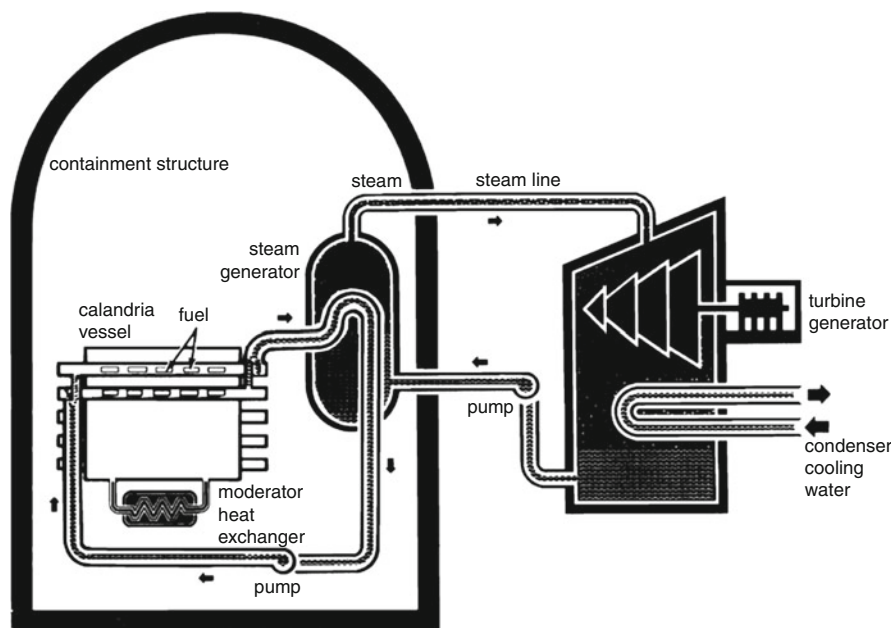
Cutaway of a representative PWR station – Standardized Nuclear Unit Power Plant System (SNUPPS) (Courtesy of Nuclear Engineering International)

The RBMK reactor (Fig. 17) consists of nearly 1,900 vertical pressure tubes – each accommodating a pair of fuel assemblies or a control rod. The pressure tubes are surrounded by an array of long, square graphite blocks – which serve as the system’s neutron moderator – set together, side-by-side vertically approximating a cylinder of roughly 12 m diameter and 7 m height. Water introduced at the bottom of the core enters pressure tubes, and then boils in the process of removing fission heat from the fuel pins. Steam is drawn from each tube through a stainless-steel pipe for use in one of two turbine generators.

The RBMK fuel assemblies (Fig. 18) each contain two stacked subassemblies of 18 zirconium-clad fuel pins of UO_2 enriched to 1.8 wt% in ^{235}U . Stacked one atop the other, a pair fills a pressure tube to facilitate the on-line refueling paradigm that adjusts reactivity and facilitates plutonium production. Various groups of control rods are moved for routine power adjustment and for full-insertion core shutdown.

Fast Breeder Reactors

The fast-neutron-spectrum breeder reactor (FBR) design concepts are predicated on breeding more new fuel than that used to sustain the neutron chain reaction (section “[Reactor Classifications](#)”). For this



Nuclear Fission Power Plants. Figure 10

Steam cycle for CANDU pressurized heavy-water reactor (Courtesy of Atomic Industrial Forum/Nuclear Energy Institute)

purpose, fissile plutonium (^{239}Pu and ^{241}Pu) and fertile ^{238}U (natural or depleted uranium with, respectively, 0.711 wt% and 2–3.5 wt% in ^{235}U) fuel with a fast-neutron spectrum are most effective. An FBR optimized for breeding, thus, can utilize uranium at least 60 times more efficiently than a normal LWR reactor. (Note that some fast-spectrum reactors – see the section “[Advanced Reactors](#)” – may not have actual breeding as a design specification.)

The liquid-metal fast breeder reactor (LMFBR) keeps neutron energy high (in the upper keV to MeV range) by using (non-moderating) liquid sodium as a coolant. The liquid sodium, although not the heaviest coolant available is not too light, has favorable heat transfer properties, and is not an excessively strong absorber of neutrons compared to other choices.

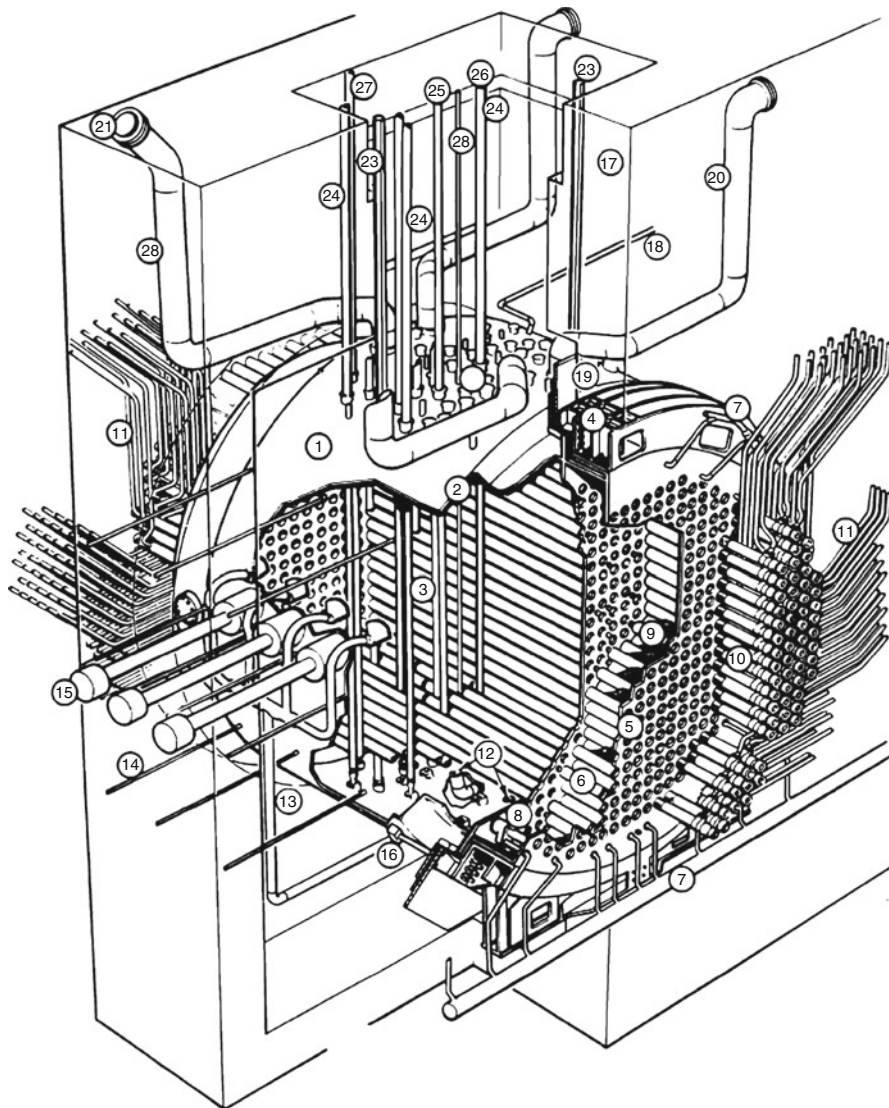
Although experimental fast breeder reactors have operated in the USA since the late 1950s, and subsequently as prototypes in Western Europe, most recent activity has been in Russia, Japan, and India.

The steam cycle has three loops (Fig. 19) with the first two of sodium and the third of water. The intermediate sodium loop is present to isolate the primary from possible direct contact with water in the steam generator. The primary sodium becomes radioactive

from neutron absorption and also can pick up fission-product radionuclides from the fuel. If this sodium were to come in contact with water, it would not only precipitate an exothermic reaction but also spread radioactive contamination.

LMFBRs have been implemented in two main configurations – loop-type and pool-type. The former is essentially as shown in Fig. 19 and similar paradigm to LWR and HTGR. The pool-type LMFBR has the reactor vessel, heat exchanger, and associated piping – entire primary loop – immersed in a large pool of sodium. The now shutdown SuperPhenix was a pool-type LMFBR of similar capacity to the current large commercial reactors. The SuperPhenix reactor vessel and layout are shown in Fig. 20.

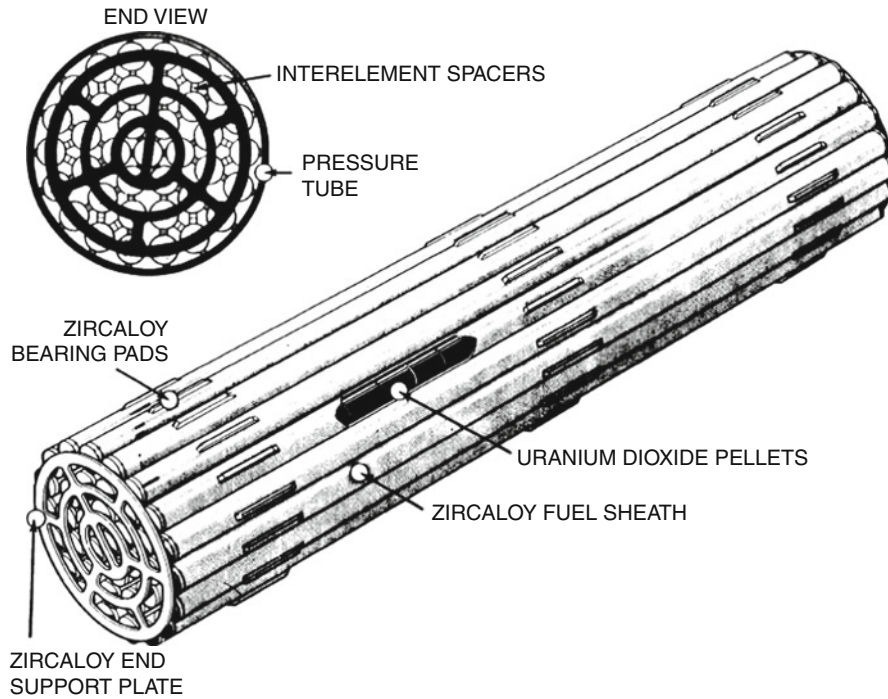
Fuel for the LMFBR consists of mixed-oxide ($\text{PuO}_2\text{-UO}_2$) fuel pellets, which typically combine about 10–30 wt% fissile plutonium with natural or depleted uranium – the remnants of the enrichment process. As shown in Fig. 21a, for a reactor other than SuperPhenix, pellets are loaded into thin stainless-steel cladding tubes, and hence into hexagonal-array sub-assemblies. Breeding is optimized by surrounding the mixed-oxide core with a *blanket* of depleted uranium. The axial blankets above and below the core are created



- | | | |
|----------------------------|-------------------------------------|-----------------------------------|
| 1 CALANDRIA | 11 FEEDER PIPES | 21 PRESSURE RELIEF DISC |
| 2 CALANDRIA SHELL | 12 MODERATOR OUTLET | 22 REACTIVITY CONTROL ROD NOZZLES |
| 3 CALANDRIA TUBES | 13 MODERATOR INLET | 23 VIEWING PORT |
| 4 EMBEDMENT RING | 14 FLUX DETECTOR | 24 SHUTOFF ROD |
| 5 FUELLING TUBESHEET | 15 ION CHAMBER | 25 ADJUSTER ROD |
| 6 END SHIELD LATTICE TUBES | 16 EARTHQUAKE RESTRAINT | 26 CONTROL ABSORBER ROD |
| 7 END SHIELD COOLING PIPES | 17 VAULT WALL | 27 ZONE CONTROL ROD |
| 8 INLET OUTLET STRAINER | 18 MODERATOR EXPANSION TO HEAD TANK | 28 VERTICAL FLUX DETECTOR |
| 9 STEEL BALL SHIELDING | 19 CURTAIN SHIELDING SLABS | |
| 10 END FITTINGS | 20 PRESSURE RELIEF PIPES | |

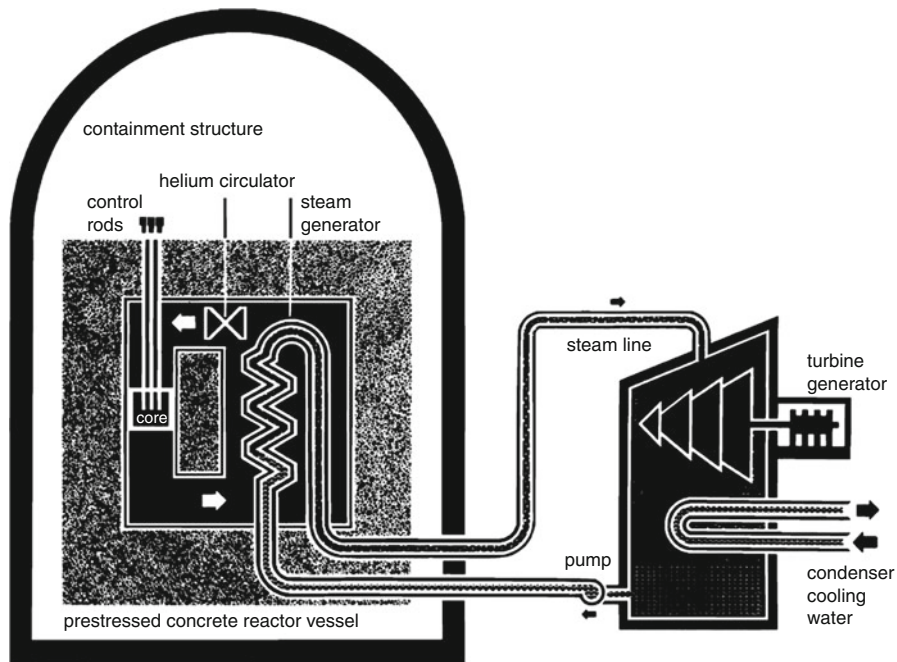
Nuclear Fission Power Plants. Figure 11

Calandria vessel and pressure tube of CANDU pressurized heavy-water reactor (Courtesy of Atomic Energy of Canada Limited)



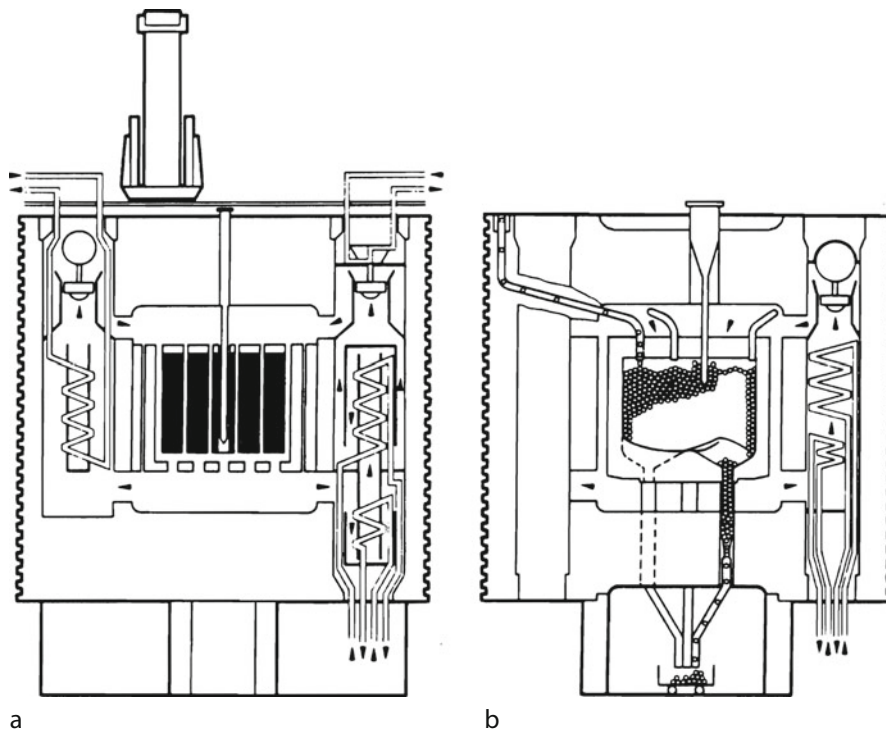
Nuclear Fission Power Plants. Figure 12

Fuel assembly for CANDU pressurized heavy-water reactor (Courtesy of Atomic Energy of Canada Limited)



Nuclear Fission Power Plants. Figure 13

Steam cycle for high-temperature gas-cooled reactor (HTGR) (Courtesy of Atomic Industrial Forum/Nuclear Energy Institute)



Nuclear Fission Power Plants. Figure 14

Prestressed concrete reactor vessels (PCRV) for (a) prismatic high-temperature gas-cooled reactor and (b) thorium high-temperature ("pebble-bed") reactor (Courtesy of Oak Ridge National Laboratory)

by loading pellets of natural or depleted uranium at either end of the core fuel pins. The surrounding radial blanket consists of separate subassemblies (Fig. 21b) of natural or depleted uranium, where the pins may be of larger diameter because the power density in the blanket is much lower than it is in the core.

LMFBR reactivity control is accomplished through use of neutron-poison control rods. Since the breeder produces more fuel than it uses, however, the core reactivity does not decrease with fissile burnup and fission-product buildup as dramatically as in the in thermal-neutron-energy converter reactors described previously. This is fortunate, because the traditional neutron poisons are not very effective for fast neutrons.

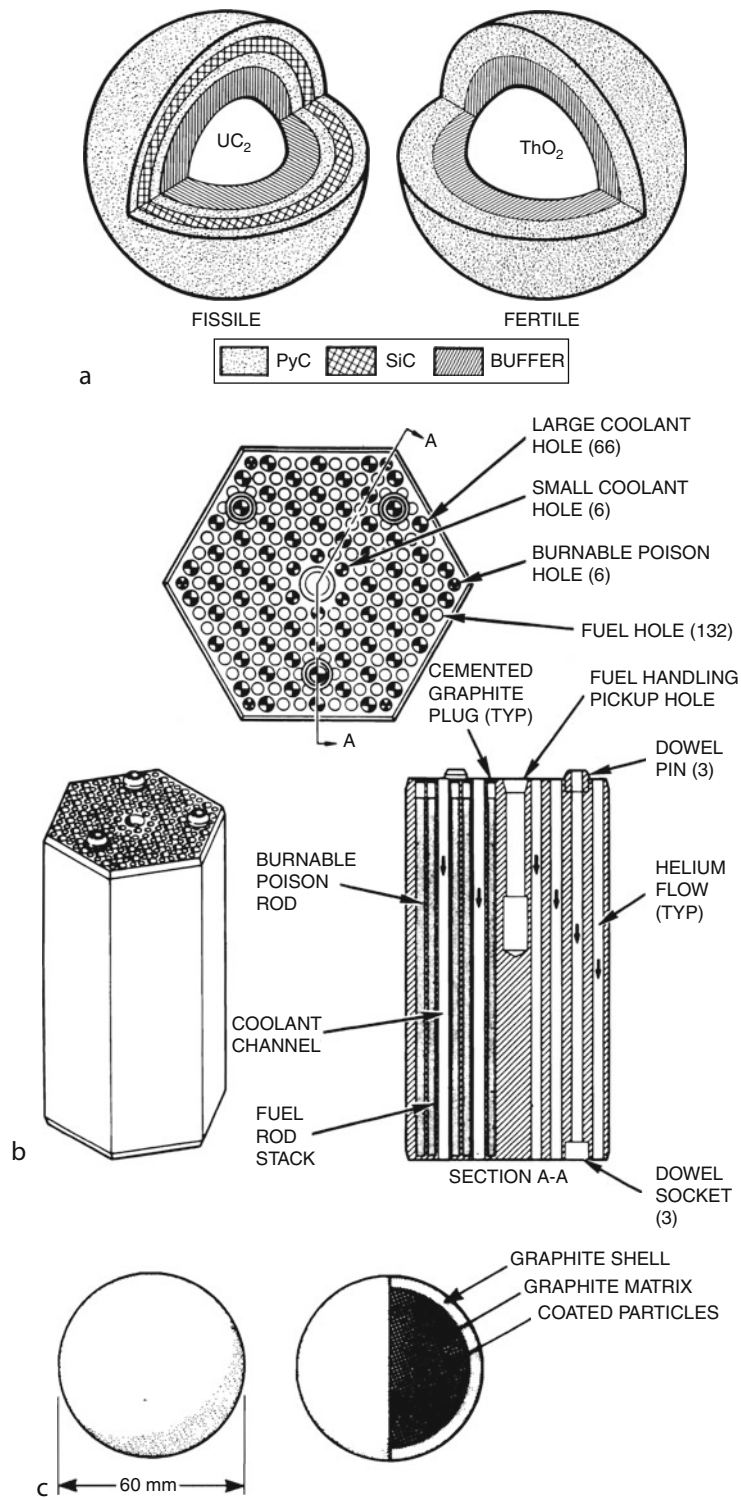
Rapid shutdown may be implemented with either of two independent sets of control rods. It is typical for one set to consist of solid neutron-poison rods, top-mounted and inserted by gravity – the traditional approach for most of the world's reactors. The second system – providing full redundancy should the first fail – often has features allowing for operation even if

core distortion has occurred, e.g., through use of oversized channels and segmented rods. Table 1 contains parameters for a large LMFBR and other reactor types. Advanced fast-neutron reactors are addressed in the section "Advanced Reactors."

Worldwide Perspective

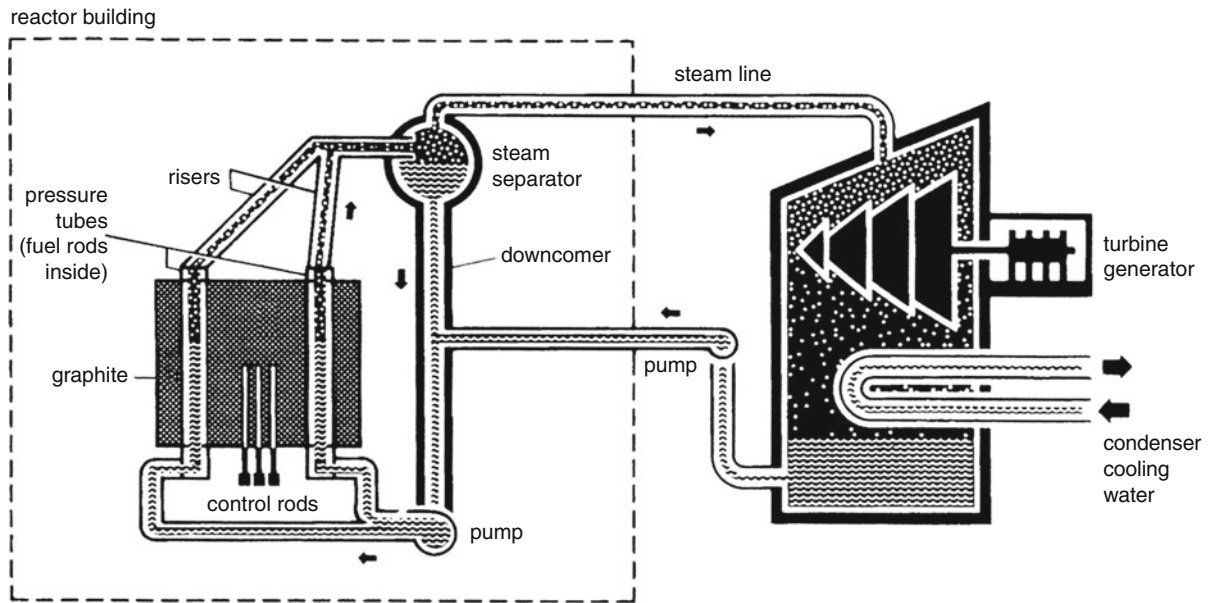
Worldwide nuclear power reactor status is shown in Table 2 as a snapshot circa 2009 of the ever-changing landscape. Reactor data are from Ref. [7], which also provides significant information on essentially each nuclear reactor ever built or even ordered.

The table also shows nuclear generation performance – as both TWe-hr generated and percentage of total national electric energy production – for 2009 and, as a comparison, a decade and a year earlier. It is apparent, e.g., that the USA and many other countries experienced a decade of stagnation at the same as the Pacific-basin nations produced impressive growth.



Nuclear Fission Power Plants. Figure 15

Fuel-assembly components for high-temperature gas-cooled reactors: (a) microspheres, (b) prismatic fuel block (Courtesy of GA Technologies), and (c) fuel sphere ("pebble") [8]



Nuclear Fission Power Plants. Figure 16

Steam cycle for pressure-tube graphite reactor (PTGR) (Courtesy of Atomic Industrial Forum/Nuclear Energy Institute)

Other Reactor Concepts

In addition to the nuclear power reactor designs described above, a wide variety of other possibilities have been built for power production or research purposes. Still others have been researched “on paper.” The major approach is to look at viable combinations of fuel, coolant, and moderator. A few examples are identified here and in the section “[Advanced Reactors](#).”

CANDU reactors could also be operated with fuel assemblies of enriched uranium, plutonium, ^{233}U /thorium, or a mixture thereof. Potential coolants other than heavy water include light water and organic liquid.

The gas-cooled fast breeder reactor (GCFBR) uses helium coolant and plutonium fuel in a concept similar to the HTGR, except with no graphite moderator. The molten-salt breeder reactor (MSBR) concept includes liquid fuel that circulates through a graphite-block core region in a closed primary loop. The fuel is processed on-line to remove fission products and ^{233}U bred from thorium in the salt.

Many other novel designs that stress enhanced safety and/or compact size also have been proposed. A wide variety of such concepts are identified in the section “[Advanced Reactors](#).”

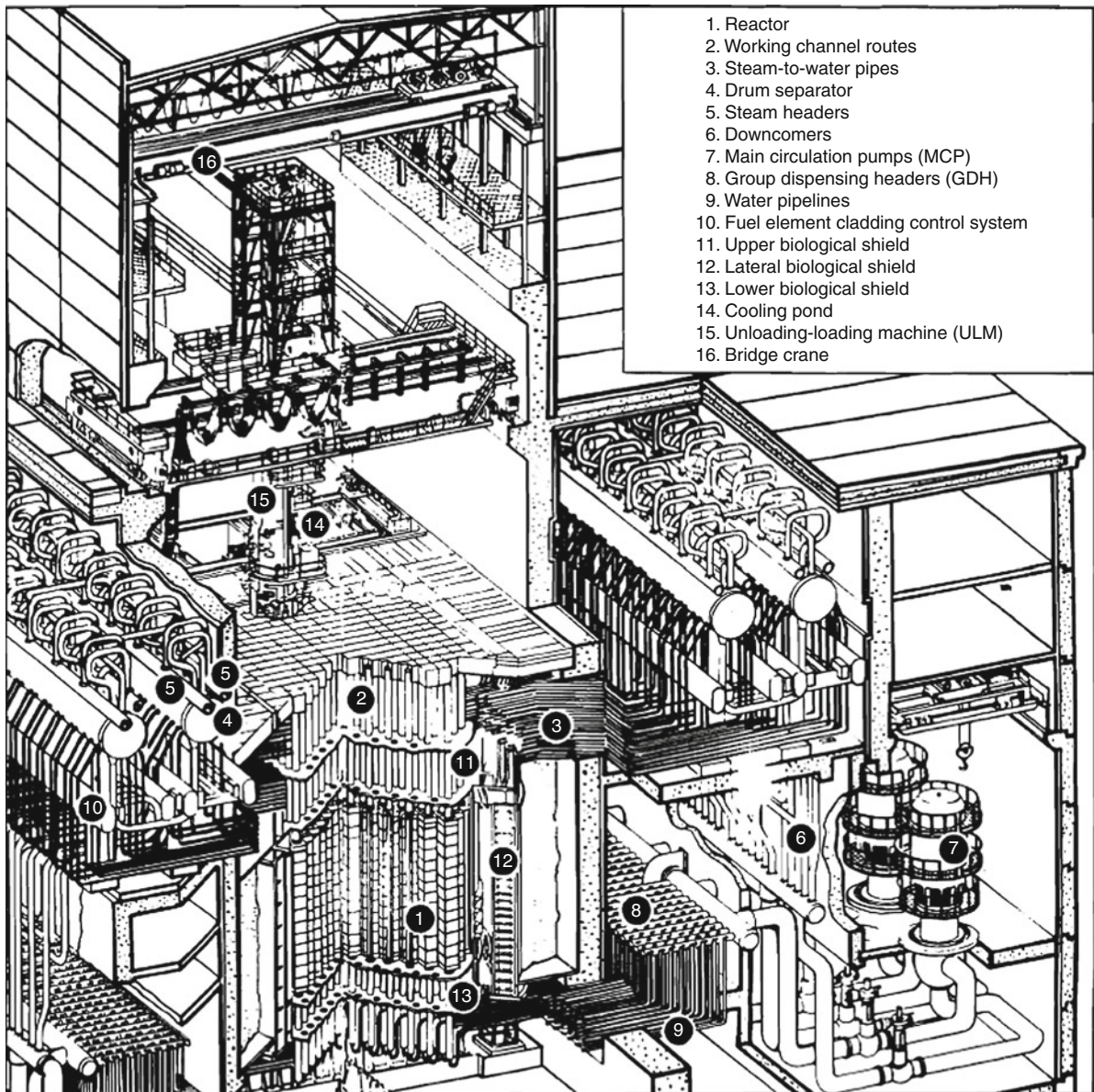
Operation, Maintenance, and Control Practices

Specific characteristics of the fission process and the design of reactors and their fuel result in unique features which affect operations, maintenance, and plant control. With over 50 years of experience for nuclear reactors in general and 30 or more years of experience with the more mature system technologies, an increasingly solid basis for good operating practices is becoming available [13]. Uniformly excellent nuclear plant availability is a key indication of this.

Operation

Routine power operations reflect characteristics of the reactor design – fuel, moderator, coolant, geometry. The BWR, with water boiling in the core, and PWR, with pressurized liquid water in the core – the two plant types dominating the world circa 2010 – show quite different operational characteristics.

Boiling-Water Reactor Startup of a boiling-water reactor begins with running the coolant pumps at minimum speed. The control rods are withdrawn from their initial fully inserted position one at a time in a preset sequence that keeps the power distribution



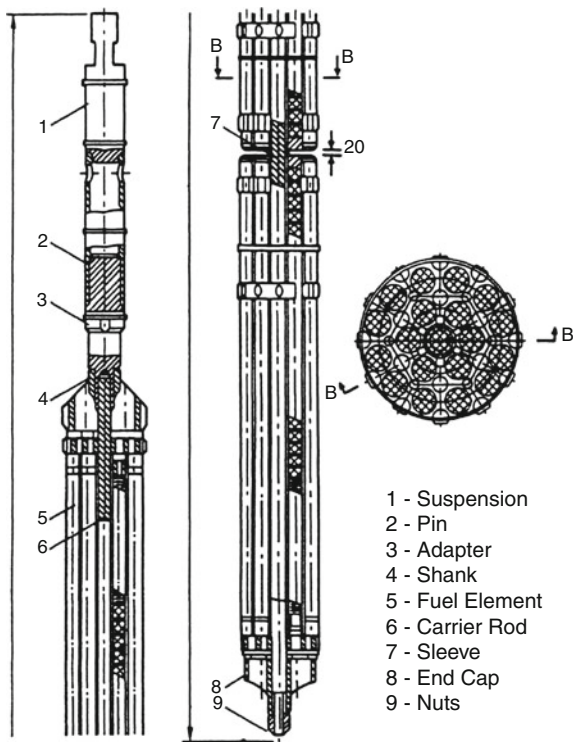
Nuclear Fission Power Plants. Figure 17

Sectional view of a Soviet pressure-tube graphite reactor (PTGR) (From [23])

reasonably uniform. The reactor is brought to critical, and then made slightly supercritical to heat the coolant slowly to the boiling point where steam formation begins to pressurize the vessel. Initially, the steam is dumped to the main condenser (Figs. 1 and 3). When operating temperature and pressure are reached and the steam flow is about 20% of full capacity, the turbine is started and the generator synchronized to the electric

power grid. Power level is increased by control-rod withdrawal, flow increase, or a combination (e.g., with flow set for 20% power, rod withdrawal takes the power to 50% and then the pumps are used to achieve full power).

Once at power, an automatic control system is used to change coolant flow for the reactor to *load follow* the turbine steam demand. The control rods are still under



Nuclear Fission Power Plants. Figure 18

Fuel assembly for a Soviet RBMK pressure-tube graphite reactor (PTGR) (From NUREG-1250 [1987])

operator control. For continued steady power operation, control-rod withdrawal or coolant flow changes can be made manually within predetermined power-to-flow limits. Control rods are withdrawn systematically to match fuel-assembly burnup.

Tripping – initiating a scram – of the control rods to shutdown a reactor causes thermal-mechanical stress. Thus, the preferred method for shutdown is essentially the reverse of the startup process. When the power drops too low for production of electricity, the steam is again dumped to the condenser. This is continued until the fission-product-decay heat load is small enough to be handled by a specifically designed residual-decay-heat-removal system.

BWR refueling is similar to that for the PWR – as described in the next subsection. One important difference is that the control rods do not need to be removed with the vessel head (because they are bottom mounted).

Pressurized-Water Reactor Startup of a pressurized-water reactor begins with alignment of support systems

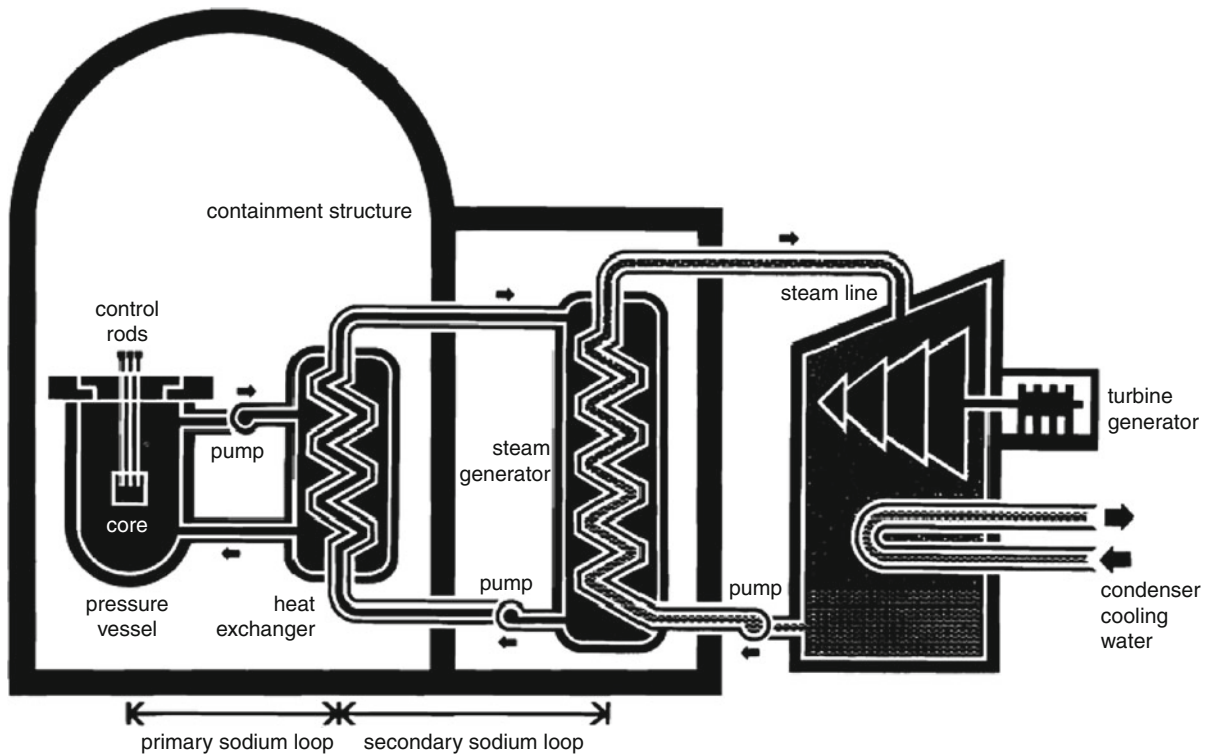
and pressurization of the primary reactor coolant system (Figs. 6 and 7). All shutdown (or “safety”) rods are “cocked” (i.e., withdrawn fully) to be available for reactor trip if necessary. Designated regulating rods remain fully inserted. The main reactor coolant pumps are started with their inherent heat generation producing a slow heat-up of the reactor coolant system that minimizes mechanical stress. As operating temperatures are approached, a “bubble” (i.e., steam volume) is drawn in the pressurizer, volume control is established, steam may be drawn to warm the steam generators, and final adjustments are made to the reactor coolant system chemistry and the soluble boron concentration. Steady withdrawal of the regulating control rods first takes the core critical and then supercritical to increase the power level at a predetermined rate. The turbine is started and the generator is synchronized to the grid. Manual control of the reactor, turbine, and steam-generator feedwater is maintained until about 15% power when transfer is made to an automatic mode for escalation to full power. In general, the turbine is matched to the reactor’s (or actually steam generator’s) steaming rate, a *reactor following* mode of operation. (Note: In practice for a PWR the reactor leads turbine, while for a BWR the reactor follows the turbine.)

During routine full-power operation, small movements of the first regulating group balance normal power fluctuations. Larger insertions change the power level.

Boric acid concentration is adjusted to match fuel burnup and allow control-rod positions to remain consistent with the power-dependent insertion limits. This concentration may be as high as 2,000 ppm (parts-per-million) in a fresh core and be reduced to 100 ppm or less by the end of the core’s lifetime.

Planned shutdown involves insertion of the rods (with additional boration if cold shutdown, e.g., for defueling, is desired). Decay heat is removed initially by one-pump or natural-circulation operation of the steam generators with the steam dumped to the condenser. When the bulk coolant temperature falls sufficiently, cooling is shifted to a dedicated residual-decay-heat-removal system and the pressurizer bubble is collapsed.

Within a day or two of reactor shutdown, the decay-heat load decreases sufficiently for reactor coolant



Nuclear Fission Power Plants. Figure 19

Steam cycle for liquid-metal fast breeder reactor (LMFBR) (Courtesy of Atomic Industrial Forum/Nuclear Energy Institute)

system depressurization and refueling. The reactor cavity is flooded with highly borated water to a height of 10 m or more so that the vessel head (with all control rods attached) and the fuel assemblies can be removed while still having enough water above them to provide radiation shielding for operating personnel. Individual spent-fuel assemblies are extracted remotely from the core and moved to a spent-fuel storage rack, usually in a separate building. Fuel not discharged is “shuffled” to new locations. Fresh assemblies are brought in from a fuel handling building.

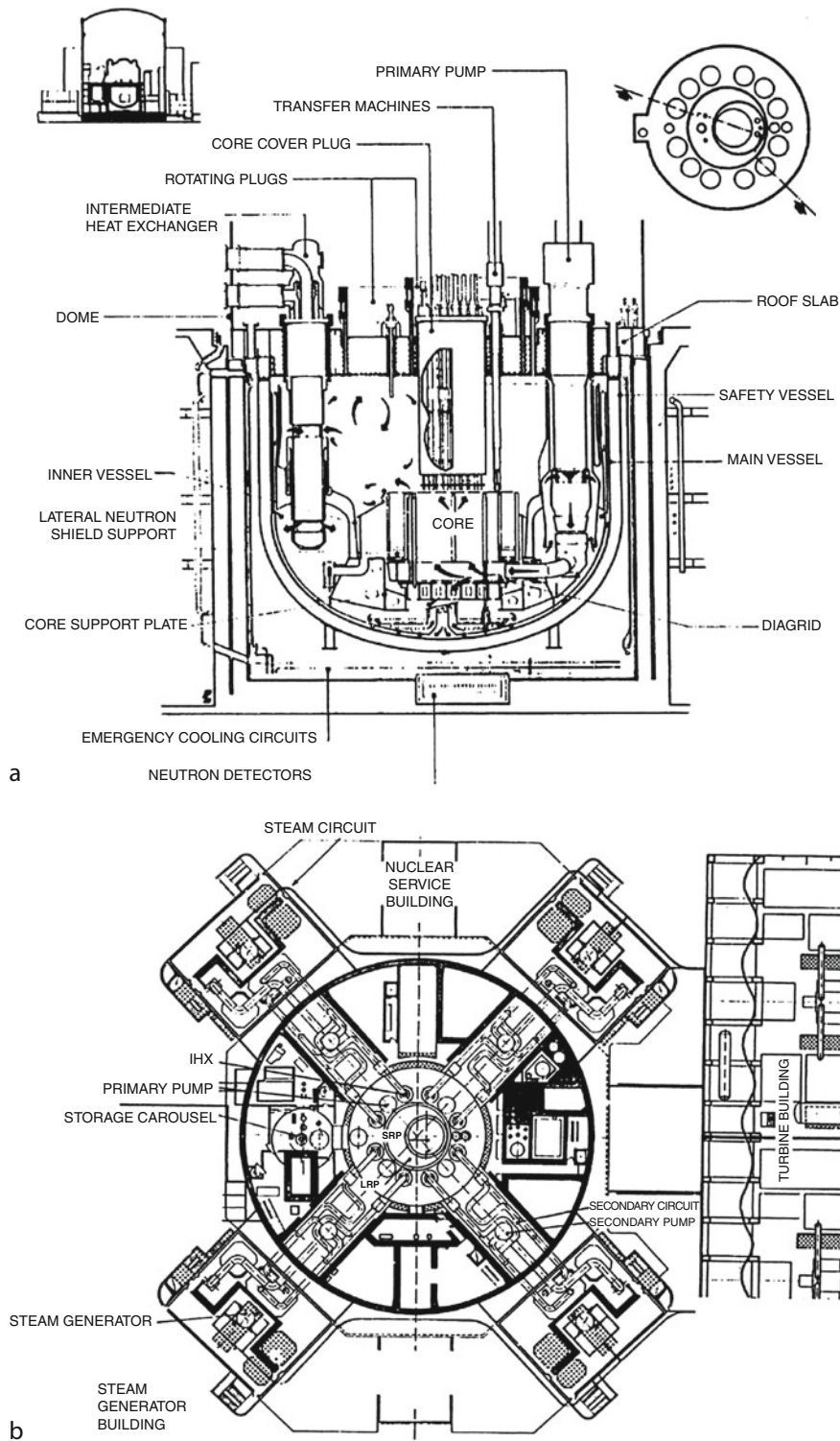
Refueling itself for an LWR requires only about 4 days. However, “refueling outages” often last up to a month or more when including cool-down and advance preparations, maintenance (corrective and preventive), minor plant modifications, vessel reassembly, and preoperational testing.

Operation Features Four key goals of operations for nuclear power plants – the first two unique and the other two common to all power plants – relate to:

(1) nuclear safety, (2) communications, (3) technical results, and (4) operating and fuel costs. Nuclear safety, i.e., prevention of accidents, underlies both technical and economic performance. Good, reliable communications with the public, staff, and company are necessary to counterbalance unfounded fears that may be associated with nuclear energy.

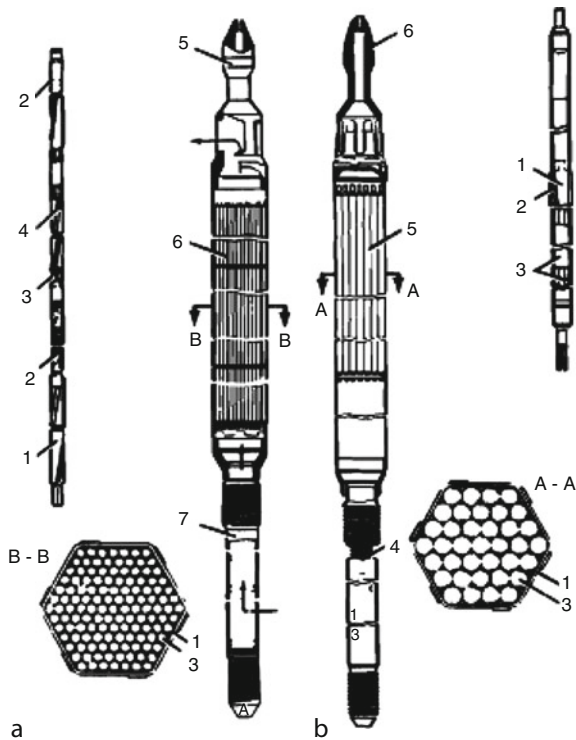
Technical results are judged according to performance indicators, an important one being availability. As for any power plant, availability is enhanced by reducing the length of planned shutdowns and the frequency of unplanned shutdowns. Because reactor fuel assemblies are added and replaced as units, adjustment of cycle length (typically between 1 and 2 years) can affect both availability and reasonable operating and fuel costs.

Where possible, nuclear units run base-loaded due to both high plant capital costs and difficulties in making rapid power-level changes (based on the behavior of certain fission products that poison the chain reaction). France, however, with its large fraction of nuclear



Nuclear Fission Power Plants. Figure 20

Basic features of the pool-type SuperPhenix (a) reactor and (b) plan-view layout (Reprinted with permission of NOVATOME, a division of Framatome)



Nuclear Fission Power Plants. Figure 21

Fuel assembly for a representative liquid-metal fast breeder reactor. (a) Fuel assembly: 1, pin cladding; 2, slugs of depleted uranium; 3, fuel pellets; 4, wire-wrapped pin; 5, fuel assembly head; 6, fuel-pin assembly; 7, stem. (b) Radial blanket assembly: 1, pin cladding; 2, wire-wrapped fin; 3, depleted uranium; 4, blanket assembly stem; 5, blanket pin assembly; 6, blanket assembly head (Courtesy of *Nuclear Engineering International*)

electricity (e.g., see [Table 2](#)) has developed technical ability to facilitate daily load following variations. The latter is also a valued attribute in a number of advanced reactors (e.g., units considered in the section “[Advanced Reactors](#).”)

The hazard of nuclear power associated with the large fission-product inventory and the potential severe accident consequences leads to special attention to operator training and to emergency response capabilities. Operators are provided with extensive technical training including substantial use of high-fidelity, plant-referenced control-room simulators which allow practice on routine, anticipated upset, and emergency operating scenarios. Emergency preparedness

includes developing procedures, establishing communications, preparing equipment and facilities, and designating an emergency response team personnel.

Maintenance

Due to the presence of fission products and other indirect sources of radioactivity from the neutron chain reaction, many important maintenance and other activities must be conducted in a manner that minimizes the hazards associated with radiation. This has led to the development of remote and other sophisticated radiological control methods. In performing hands-on work, it is challenging to hold radiation exposures of personnel within specified limits on quarterly, annual, and lifetime doses. Overall, however, the requirement that all exposures be kept “as low as reasonably achievable” (ALARA) has resulted in extensive preplanning of all radiological work activities.

Gaseous and liquid wastes are held for as long as practicable to allow maximum radioactive decay. Then, they may be diluted and dispersed to the environment so long as applicable limits for liquid and airborne discharges are not exceeded. Otherwise, solidification, containment, and disposal are required.

Nuclear reactors, like all power plants, require corrective and preventive maintenance that balance the conflicting goals of maximizing short-term profits by deferring work and reducing intervals between regular maintenance activities to assure top performance. The amount of corrective maintenance depends primarily on the initial quality of components and systems, control of operating conditions, and the effectiveness of the preventive maintenance program.

Preventive maintenance for the reactor coolant system (including reactor vessel, reactor coolant pumps, control rod drive mechanisms, instrumentation, and in applicable systems pressurizer and steam generators) must account for neutron irradiation effects on components and systems. The nuclear industry is using methods of reliability-centered maintenance increasingly as sufficient operating experience and data are collected to provide useful statistical reference. In France, where a single utility operates a sizeable number of each of several standardized units, maintenance is considered a lifetime project with integrated goals to guarantee safety, ensure efficiency, assess impact of

Nuclear Fission Power Plants. Table 2 Worldwide nuclear generating capacity by reactor type^a and summary of reactor performance

Country ^c	Reactors operable		Reactors under construction		Reactors under contract and [planned] ^{b,c}		1998 reactor performance ^d		2009 reactor performance ^e	
	No.	MWe	No.	MWe	No.	MWe	TWe-hr	% of Electric	TWe-hr	% of Electric
Argentina							6.9	10.0	7.6	7.0
PHWR	2	1,005	1	745						
Armenia							1.42	24.7	2.3	45
VVER	1	408			1	1,060				
Belgium							43.9	55.2	45	51.7
PWR	7	6,192								
Brazil							3.3	1.1	12.2	3.0
PWR	2	2,007	1	1,309						
Bulgaria							15.5	41.5	14.2	35.9
VVER	2	2,000								
Canada							67.5	12.4	85.3	14.8
PHWR	17	12,969			[5]	4,800				
China							13.5	1.2	65.7	1.9
PWR	7	5,558	21	23,000	5/[42]					
VVER	2	2,000			[4]					
PHWR	2	1,500								
HTGR	1	2	1	195						
FBR					[3]					
	12		22	23,195	5/[49]					
Cuba										
PWR			2	880 suspend						
Czech Republic							28.5	20.5	25.7	33.8
VVER	6	3,892			[2]					
Finland							21.0	27.4	22.6	32.9
VVER	2	1,020								
BWR	2	1,780								
EPR PWR			1	1,600						
Unspecified					[3]					
TOTAL	4	2,800	1	1,600	[3]					
France							368.4	75.8	391.7	75.2

Nuclear Fission Power Plants. Table 2 (Continued)

Country ^c	Reactors operable		Reactors under construction		Reactors under contract and [planned] ^{b,c}		1998 reactor performance ^d		2009 reactor performance ^e	
	No.	MWe	No.	MWe	No.	MWe	TWe-hr	% of Electric	TWe-hr	% of Electric
PWR	58	63,134	1	1,650	[1]	[1 650]				
Germany							145.2	28.3	127.7	26.1
PWR	11	14,783								
BWR	4	4,526								
	12	19,309								
Hungary					[2]		13.1	35.6	14.3	43
VVER	4	2,000								
PWR										
India							10.2	2.5	14.8	2.2
PWR					[4]					
EPR PWR					[6]					
BWR	2	320			[8]					
PHWR	16	4,140			[4]					
VVER			2	2,000	[4]					
FBR					[1]					
	18	4,460			[27]					
Iran										
VVER			1	1,000						
Japan							306.9	35.9	263.1	28.9
PWR	24	20,278			2	3,076				
BWR	25	23,221	1	1,373	[1][3]					
FBR					[1]	660				
ABWR					2 [6][1]	2,746				
APWR					2	1,590				
	49	43,499	1	1,373	6 [12]					
Kazakhstan							0.1	0.2		
VVER					[3]	1,920				
(Korea, North)										
PWR					[2] cancel	1,000				
Korea, South							85.2	41.4	141.1	34.8
PWR	16	15,643	6	6,800	[2]	2,800				

Nuclear Fission Power Plants. Table 2 (Continued)

Country ^c	Reactors operable		Reactors under construction		Reactors under contract and [planned] ^{b,c}		1998 reactor performance ^d		2009 reactor performance ^e	
	No.	MWe	No.	MWe	No.	MWe	TWe-hr	% of Electric	TWe-hr	% of Electric
PHWR	3	2,189								
	19	17,832								
Lithuania							12.3	77.2	10	76.2
RBMK	2 s/d	2,600								
Unspecified					[1]					
Mexico							8.8	5.4	10.1	4.8
BWR	2	1,364								
Netherlands							3.6	4.1	4.0	3.7
PWR	1	515								
Unspecified					[1]	1,000–1,600				
Pakistan							0.3	0.7	2.6	2.7
PHWR	1	137								
PWR	1	325	1	325						
Unspecified					[1]	600				
	2	462								
Romania							4.9	10.4	10.8	20.6
PHWR	2	1,411			[2]	1,412				
Russian Federation							95.4	13.1	152.8	17.8
VVER	16	12,594	4	4,600	4 [6][19]	4,600				
RBMK	11	11,000	1	1,000						
LWGR (EGP)	4	48			[3]					
KLT (Floating)			2	70						
FBR	1	600	1	880						
BWR (heat)					[4]					
VBER (PWR)					[6]					
	32	14,242	8	6,550	4 [38]	4,600				
Slovakia							11.4	43.8	13.1	53.5
VVER	4	1,896	2	880						
Slovenia							4.8	38.3	5.5	37.9
PWR	1	730								

Nuclear Fission Power Plants. Table 2 (Continued)

Country ^c	Reactors operable		Reactors under construction		Reactors under contract and [planned] ^{b,c}		1998 reactor performance ^d		2009 reactor performance ^e	
	No.	MWe	No.	MWe	No.	MWe	TWe-hr	% of Electric	TWe-hr	% of Electric
South Africa							13.6	7.25	11.6	4.8
PWR	2	1,880			[1]	110				
HTGR										
Spain							56.7	31.7	50.6	17.5
PWR	6	5,912								
BWR	2	6,177								
	8	12,089								
Sweden							70.0	45.8	50.0	34.7
PWR	3	3,002								
BWR	7	6,559								
		9,561								
Switzerland							24.4	41.1	26.3	39.5
PWR	3	1,795								
BWR	2	1,610								
		3,405								
Taiwan									(39.3)	(17.1)
PWR	2	1,914								
BWR	4	3,276	2	2,700						
	6	5,190								
Turkey										
VVER					[2]					
Ukraine							70.6	45.4	77.9	48.6
VVER	15	13,835	2	2,000						
United Arab Emirates										
PWR					2	2,000				
UK							91.1	27.1	62.9	17.9
PWR	1	1,250								
Magnox GCR	4	1,540								
AGR GCR	14	9,112								
EPR PWR					[4]	6,600				
Unspecified					[4]					
	19	11,902			[8]					

Nuclear Fission Power Plants. Table 2 (Continued)

Country ^c	Reactors operable		Reactors under construction		Reactors under contract and [planned] ^{b,c}		1998 reactor performance ^d		2009 reactor performance ^e	
	No.	MWe	No.	MWe	No.	MWe	TWe-hr	% of Electric	TWe-hr	% of Electric
USA							673.7	18.7	798.7	20.2
PWR	69	70,388	1	1,218						
BWR	35	35,903								
AP1000					2 [12]	2,500				
ESBWR					[2]					
ABWR					[2]					
US-EPR PWR					[6]					
US-APWR					[2]					
Unspecified					[6]					
	104	106,291			2 [30]	2,500				
TOTAL ^f										
PWR	214	215,306	31	34,302	9 + [53]					
EPR PWR					[10]					
US-EPR PWR					[6]					
APWR					[2]					
US-APWR					[2]					
AP1000					[12]					
VVER	52	39,645	11	10,480	7 + [40]					
VBER					[6]					
BWR	84	84,736	3	4,073	[12]					
ABWR					2 + [9]					
ESBWR					[2]					
BWR (Heat)					[4]					
PHWR	43	23,351	1	745	[11]					
HTGR	1	2	1	195	[1]					
Magnox	4	1,540								
AGR	14	9,112								
RBMK	11	11,000	1	1,000						
LWGR (EGP)	4	48			[3]					
FBR	1	600	1	800	[5]					

Nuclear Fission Power Plants. Table 2 (Continued)

Country ^c	Reactors operable		Reactors under construction		Reactors under contract and [planned] ^{b,c}		1998 reactor performance ^d		2009 reactor performance ^e	
	No.	MWe	No.	MWe	No.	MWe	TWe-hr	% of Electric	TWe-hr	% of Electric
KLT-Float			2	70						
Unspecified					[17]					
TOTAL	414	385,340	51	51,665	18 + [139]					

^aReactor data current through 2009. Source: *2010 World Nuclear Industry Handbook*, from the publishers of Nuclear Engineering International magazine, www.neimagazine.com. (Note: Reference also contains details on essentially every reactor built or planned)

^bThe source document includes several categories of planned reactors, essentially ranging from very firm to highly speculative. As used in this table, the author's designation "[Planned]" excludes the more speculative categories

^cWorld Nuclear Association reported reactors also "planned" and/or "proposed" by Bangladesh, Belarus, Egypt, Indonesia, Israel, Italy, Jordan, Malaysia, Poland, and Thailand – <http://www.world-nuclear.org/info/reactors.html> – August 1, 2010 (updated periodically)

^dFrom *Nuclear Engineering International World Nuclear Industry Handbook 1999*, November 1998

^eFrom World Nuclear Association – <http://www.world-nuclear.org/info/reactors.html> – August 1, 2010 (updated periodically)

^fNot all countries reporting status distinguished between traditional reactors (e.g., PWR or BWR) and advanced versions (e.g., APWR or ESBWR) thereof

changes in operating conditions, minimize maintenance cost increases, and establish an optimum decommissioning policy (i.e., definition of the end of a reactor's safe and economic lifetime).

Plant Control Systems

Plant instrumentation and control systems are divided among those for normal operation, regulation of parameters through analog or digital technologies, and protection of personnel and equipment by reactor trip and safety feature actuation. Examples of control requirements in the first category may be inferred from discussions earlier in this subsection of operation of boiling- and pressurized-water systems. Specific safety features are described in the next subsection.

Reactors in the USA have tended to continue use of older analog systems for routine reactor control, lacking incentive – and, in some sense, having a *disincentive* associated with time and costs associated with being among the first to develop, test, and license a process – to upgrade to state-of-the-art digital systems. However, significant government-sponsored research and development support in the first decade of the twenty-first century offers the prospect of near-term progress.

Data acquisition and display technology, by contrast, has been upgraded substantially, especially in the aftermath of the accident at the Three Mile Island, Unit 2 reactor. Should the control room become uninhabitable, a remote shutdown panel will allow operators to implement the emergency response functions necessary for safe shutdown of the plant.

France, with its large, dynamic development program, has upgraded controls in early units and developed new systems. Their current and near-future reactor generations have particularly sophisticated reactor operating consoles with state-of-the-art digital systems for control, data acquisition, operator support, and emergency response. Control systems facilitate process monitoring, exchange of information, and self-monitoring and self-diagnosis. Automatic protection systems are designed for reliability with redundancy and diversity (e.g., one of which includes four redundant data acquisition and processing units coupled with two independent safeguard logic units for automatic actuation of the engineered safety features) and include testing and simulation capabilities. Data processing systems are designed to give operators full information on the present state of the reactor, monitor data, store event scenarios and other data changes in memory, and perform calculations.

Advanced reactor controls may facilitate functions such as automatic turbine load shedding, steam level control, (nonreactor, non-safety-related) and trip avoidance – and the consequent thermal-mechanical disruption, and postaccident monitoring and control.

Nuclear Reactor Safety Features

An operating nuclear power reactor accumulates an enormous inventory of radioactive products. Each reactor system has been designed with a multiple-barrier approach to retention of this radioactivity [8]. The first three of these barriers – fuel pellet, metal cladding, and reactor-primary-system – were described in the section “[Reactor Systems](#).” The fourth barrier is a containment building supported by safety systems.

A “defense-in-depth” design approach to maintaining the effectiveness of barriers seeks first of all to prevent accidents. But absent this, protective actions – identification and correction – and mitigative actions – long-term response to and control of consequences are provided. Safety analyses address the variety of reactor-related energy sources and their roles in operational abnormalities and in design-basis and more severe accidents. The results of the analyses also define requirements for the engineered safety systems.

Energy Sources

Five distinct energy sources may contribute to an accident in a nuclear power reactor:

1. Stored energy – fuel, coolant, and structures store thermal energy at all times during reactor power operation; redistribution may result in immediate damage and/or prolonged problems.
2. Nuclear transients – positive insertion of reactivity, from configuration or material changes, may add a transient energy source resulting in from an increased power level to a large power pulse.
3. Decay heat – heat from fission-product decay, as high as 7.5% of operating power at the time of shutdown from a lengthy run, dies out slowly, and can be a large energy contribution after dissipation of stored energy and shutdown of the neutron chain reaction; with inadequate heat removal, it may be sufficient to cause fuel melting or other damage.
4. Chemical reactions – fuel, cladding, and coolant materials selected to be essentially unreactive with each other under normal operating conditions at elevated temperatures that may be produced by nuclear transients and/or decay heat may experience energetic chemical reactions.
5. External events – natural (floods, hurricanes, tornadoes, and earthquakes) and man-made (aircraft impacts and industrial explosions) events external to the reactor system have the potential to initiate or otherwise contribute to accidents.

The magnitude and timing of contributions from each category are important to accident progression. Understanding each is necessary to accident prevention and mitigation.

Nearly 98% of all radioactive products are retained by the fuel assemblies when sufficient cooling is provided to prevent fuel melting. Thus, major objectives of nuclear reactor operation and safety are to provide adequate heat removal and control of the energy released in the system to prevent overheating and, in the most severe case, melting of the fuel.

Design-Basis Accidents

Design-basis accidents involve the postulated failure of one or more important systems and an analysis based on conservative assumptions (e.g., pessimistic estimates of fission-product release). The radiological consequences must be shown to be within preestablished limits. In this sense, the accidents serve as the *basis* for assessing the overall acceptability of a particular reactor *design*. Design-basis accidents for light-water reactors (and, by extension, other reactor types) are often classified according to the following general characteristics:

1. Overcooling – excessive heat removal from steam withdrawal, perhaps through a steam-line break or steam-generator overfeeding (PWR and other multi-loop systems).
2. Undercooling – inadequate heat removal occurring when steam flow is decreased, e.g., following a turbine trip or reduction in feedwater flow; a complete *loss of heat-sink accident* (LOHA) – up to and including total loss of heat-removal capability – is the most extreme case of undercooling.

3. Overfilling – increase in reactor coolant inventory, e.g., from malfunction of the volume control system or inadvertent emergency core cooling system (ECCS) – see the section “[Safety Systems](#)” – actuation during power operation.
4. Loss of flow (accident (LOFA)) – decrease in reactor coolant system (RCS) flow rate, e.g., follow failure of main reactor coolant pump(s).
5. Loss of coolant (accident (LOCA)) – decrease in reactor coolant inventory resulting from breaches in the reactor coolant pressure boundary; the “classic” LOCA involves rupture of major primary piping, but *steam-generator tube rupture* and inadvertent opening of a relief or safety valve – the proximate cause of the reactor accident at Three Mile Island – are also included.
6. Reactivity-core reactivity and power distribution anomalies – unplanned control-rod withdrawal, other maloperation, and ejection (in a PWR) or drop (in a BWR) of one or more control-rod assemblies; also increased BWR coolant flow rate, and decreased PWR boron concentration, or misleading of a fuel assembly.
7. Anticipated transient without scram (ATWS) – failure to trip (or shutdown) the reactor following a transient such as inadvertent control-rod withdrawal, turbine trip, or loss of feedwater.
8. Spent-fuel and waste system-radioactivity release from a spent-fuel assembly or reactor subsystem or component – events outside of the reactor-primary system.
9. External events-natural or human-caused events that can effect plant operating and safety systems – sequence initiated by an “external event” as identified in the section “[Energy Sources](#).”

More severe accidents, sometimes referred to as beyond-design-basis accidents, where multiple safety systems fail to function (or, as in the Chernobyl accident, are disabled), are sometimes considered to evaluate overall responses.

Safety Systems

Safety systems for accident mitigation are highly design-dependent (e.g., based on design and operational differences among fuel form: coolant – water, gas, liquid metal; neutron energy; and other features.

The common goals, however, are to prevent overheating, fuel melting and other damage, and subsequent large-scale dispersal of fission products. Reliability is enhanced through redundancy in subsystem function and location.

Reactor safety systems – for the dominant loss-of-coolant accident, but with more general applicability as well – may be classified according to function as:

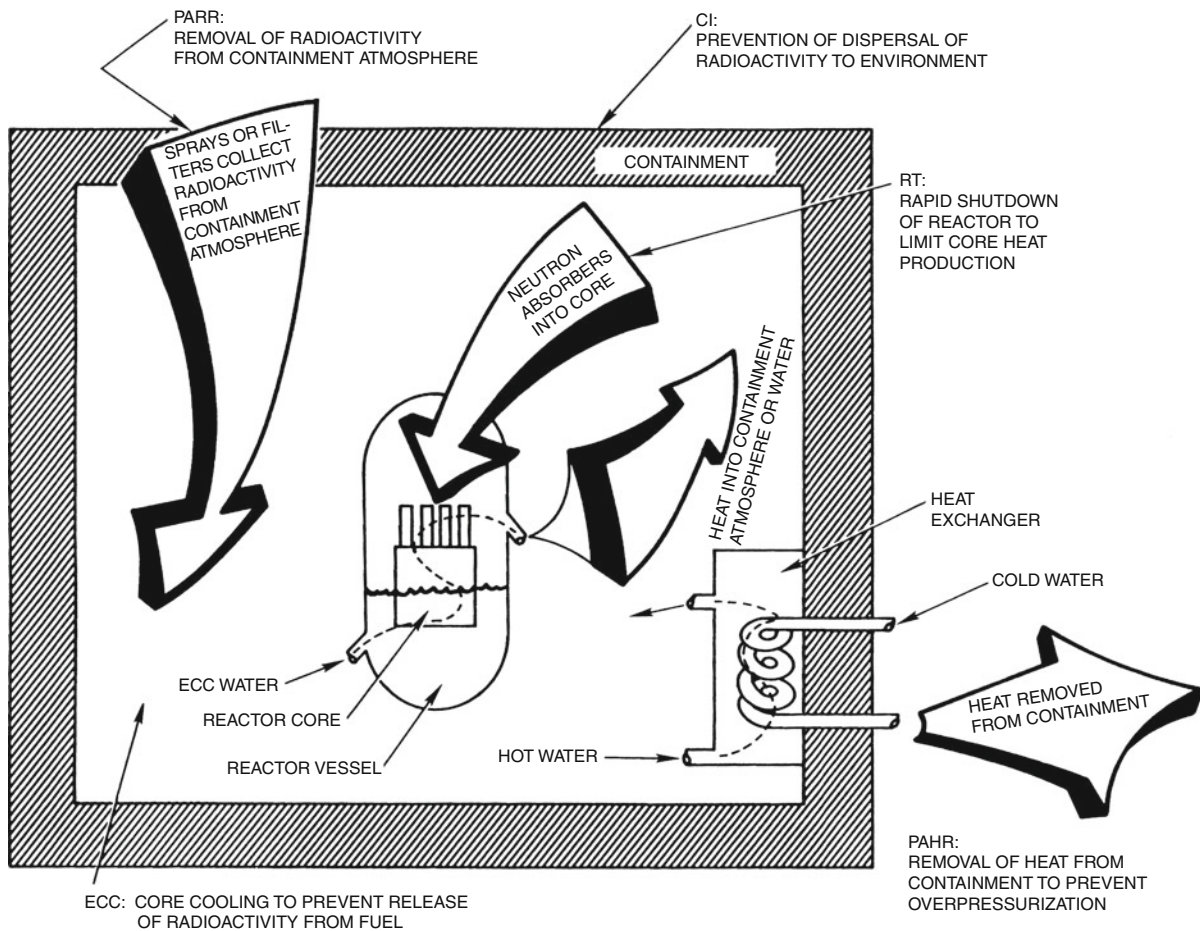
1. Reactor trip (RT)
2. Emergency core cooling (ECC)
3. Postaccident heat removal (PAHR)
4. Postaccident radioactivity removal (PARR)
5. Containment integrity (CI)

Their basic functions are summarized by [Fig. 22](#) for light-water reactors. The same basic functions apply to all reactor systems, even if in somewhat different form.

Reactor trip for each of the reactor types described in the section “[Reactor Systems](#)” includes neutron poison control rods, which can be inserted rapidly into the fuel core to shut down the fission chain reaction. These rods are supplemented by alternative shutdown means including injection of soluble boric acid poison (LWRs); moderator dump (CANDU-PHWR); reserve shutdown spheres (HTGR); and a redundant, independent set of rods (LMFBR).

Emergency core cooling for the light-water, and most other water-cooled, reactors is implemented by injection of borated water – the same medium that provides for emergency reactor trip – into the coolant-starved core region following a LOCA event. Multiple trains of separate systems typically can inject water at high, intermediate, or low pressure to coincide with various needs during the time-history and/or magnitude of the event. The recirculation of coolant that collects in the reactor building sump provides a long-term coolant supply after the initial inventories have been exhausted.

The CANDU-PHWR system also has a unique safety-related design feature with two separate flow circuits configured so that adjacent pressure tubes ([Fig. 11](#)) receive coolant from a different source. Thus, a break in a specific pressure tube can affect no more than half of the fuel and no large core volumes are left uncooled. Additionally, the separate moderator volume provided by the calandria vessel provides a huge heat sink for energy removal.



Nuclear Fission Power Plants. Figure 22

Functions of safety systems for light-water reactors (Adapted from WASH-1400 [14], courtesy of US Nuclear Regulatory Commission)

Emergency cooling for the HTGR design is predicated largely on the very high heat capacity of the graphite fuel, supported by auxiliary helium circulation and containment afforded by a steel-lined prestressed concrete reactor vessel (PCRVR). The LMFBR, by contrast with the water-cooled reactors, has low-pressure, high-boiling liquid-sodium coolant which would not be voided even during a serious primary-system breach. Thus, long-term, natural-circulation cooling capability is uniquely available. Postaccident heat removal has two components – coolant temperature reduction and containment-building pressure control. The first is accomplished through heat exchangers for ECC water recirculation in the water reactors. For the gas- and liquid-sodium-cooled

reactors, continued use of the steam generators can serve a similar function in the primary coolant loops. Pressure control may be accomplished with containment-atmosphere coolers. In water reactors this may be supplemented with steam-condensing water sprays.

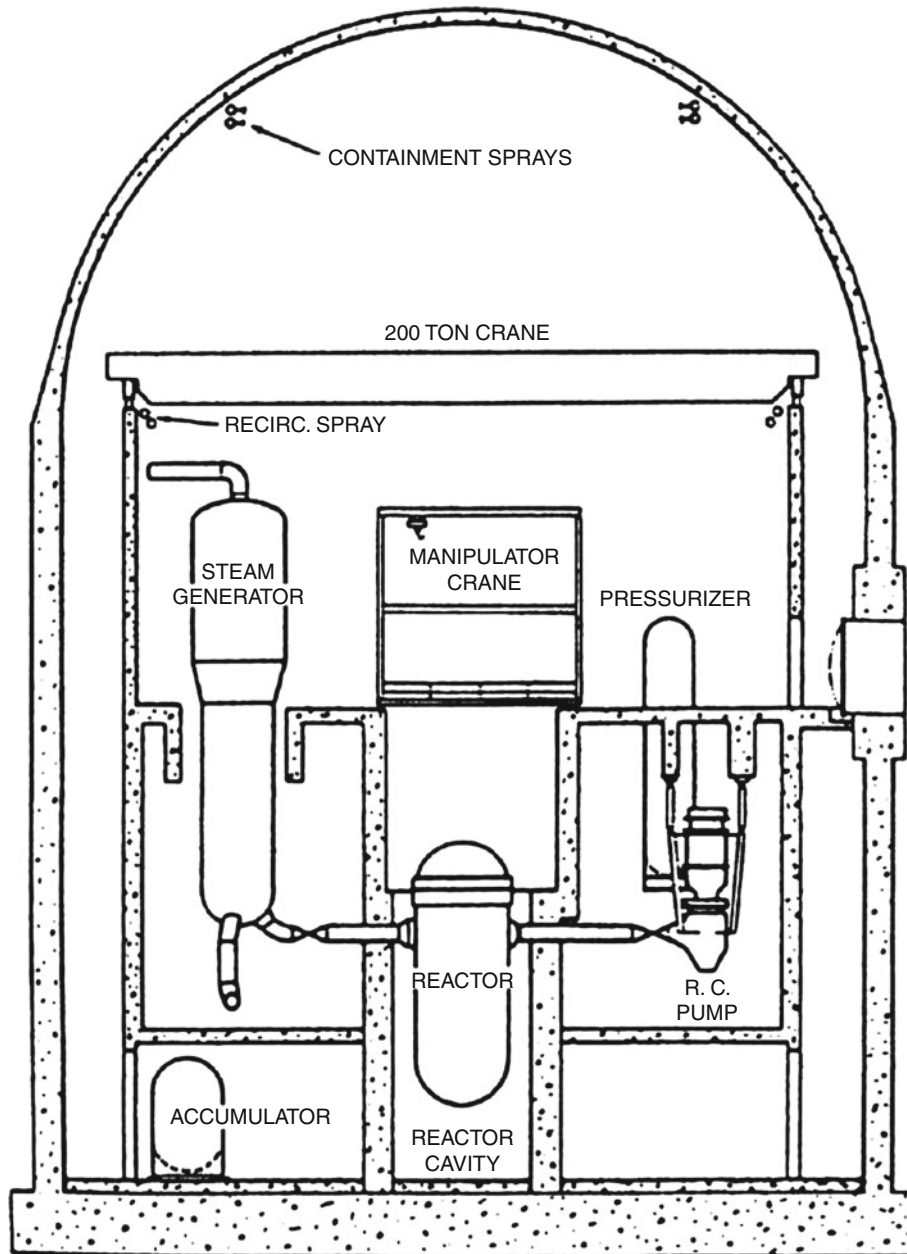
Postaccident radioactivity removal of chemically active iodine and aerosol/particulate constituents is accomplished with filtration. The nonreactive noble-gas constituents can only be contained – or released in a controlled manner.

Water-cooled reactors also have provision for containment sprays to remove radioactivity. Although the water sprays used for pressure reduction naturally remove much of the chemically reactive radioactive

material, additives such as sodium hydroxide or thio-sulfate increase removal, especially of elemental iodine.

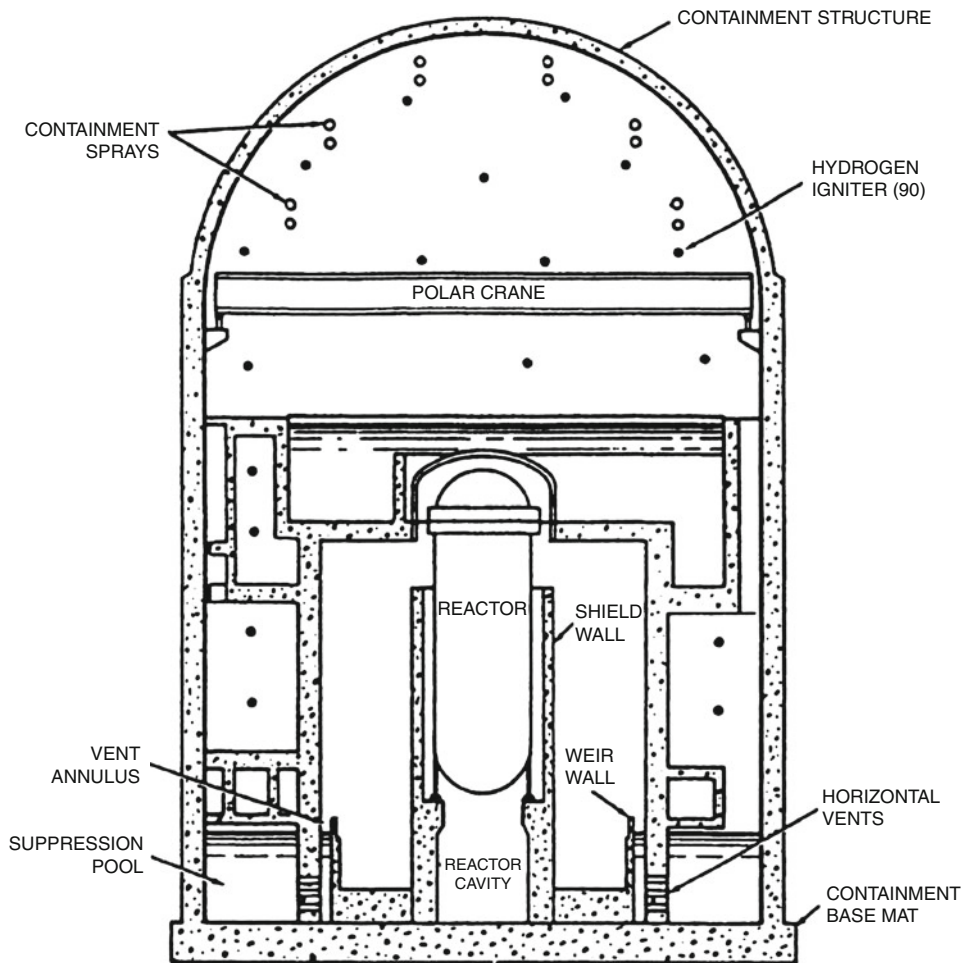
Containment integrity is the last line of defense against fission-product release. Common features of water reactor containments are a leak-tight steel liner

surrounded by thick reinforced concrete, including, e.g., that for the pressurized-water and boiling-water reactors in Figs. 23 and 24, respectively. The structure of the HTGR's prestressed concrete vessel (Fig. 14) serves a similar purpose.



Nuclear Fission Power Plants. Figure 23

Containment structure for a representative pressurized-water reactor (From NUREG-1150 [15], courtesy of US Nuclear Regulatory Commission)



Nuclear Fission Power Plants. Figure 24

Containment structure for a representative boiling-water reactor (From NUREG-1150 [15], courtesy of US Nuclear Regulatory Commission)

Another major element of containment integrity is the ability to isolate penetrations using remotely operated valves or other means. These typically actuate on the predetermined indication of excessive pressure, radiation level, or other related parameters.

Integrity of even the most robust containment, however, depends on having all – or at least a reasonable subset – of the other safety features function as intended. Otherwise there can be excessive pressure buildup and/or severe fuel melting with subsequent dire consequences. The complete lack of a leak-tight containment structure at the Chernobyl PTGR (Fig. 17) was a major contributor to the consequences of the 1986 accident there.

Defense-in-Depth

The previous descriptions of the defense-in-depth approach to nuclear safety have a strong technical emphasis. However, nuclear safety also depends on preventive measures, only a few of which include testing and inspection, safety assessment, deficiency analysis and correction, training, and quality assurance. Protection and mitigation require transient operating guidelines and procedures and detailed emergency planning.

Increasingly important is free and open international exchange of knowledge and experience with proper feedback to reactor design, maintenance, and

operation – the need for which was highlighted dramatically by the TMI-2, Chernobyl-4, and other accidents. Regulatory controls and voluntary peer oversight – e.g., by the US Institute for Nuclear Power Operations (INPO) and the World Organization of Nuclear Operators (WANO) – provide independent verification of safety aspects, but do not replace responsibility of the owner/operators.

Nuclear-Power Regulation

The inherent hazard associated with the radioactive material in reactor systems has led to the nuclear industry being the most regulated in the world [8, 13, 16]. Regulatory bodies include: the US Nuclear Regulatory Commission (NRC); French Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA) – up until late 2009, Commissariat à l'Énergie Atomique, now with added regulation of alternative energies; Canadian Nuclear Safety Commission (CNSC) – known as Atomic Energy Control Board (AECB) prior to 2000; and the UK Nuclear Installations Inspectorate (NII). Every other nation that operates nuclear-fission reactors also has established a regulatory authority. Requirements for reactor design and operation are established and their implementation is evaluated and monitored. Regulations with the force of law, licenses, or other methods are developed for reactor operation by such bodies, often with significant political and/or public input to the process.

The somewhat recent advent of advanced nuclear power reactors (section “[Advanced Reactors](#)”) has led to changes in the NRC licensing process. Newly instituted generic certifications are valid for 15 years. Using an exhaustive public process, safety issues within the scope of the certified designs are fully resolved and, hence, will not be open to legal challenge during licensing for specific plants. US utilities will be able to obtain a single NRC reactor license to both construct and operate – as opposed to separate ones previously – before construction begins. A number of reactor types have received design certification [2].

Overall, certification of designs is on a national basis, and is safety based. In Europe, there are moves toward harmonized requirements for licensing. Reactors there may also be certified according to

compliance with European Utilities Requirements (EUR) [2].

Quality assurance (QA) requirements focus on methods and procedures to assure proper design, construction, and operation of safety-related components and subsystems. Other requirements address issues from acceptable personnel radiation exposures to analysis methods and acceptance criteria for consequences of design-basis accidents and, in some cases beyond-design-basis accidents.

The accident at the Three Mile Island Unit 2 (TMI-2) PWR in 1979 led to seminal changes in the regulatory process – and nuclear industry as a whole – in the USA and elsewhere in the world. Some of these changes relate to the design, quality assurance, and inspection of modifications to plant safety systems; development and use of preapproved procedures for operation, maintenance, and other activities; administration, including staffing, training, and documentation; emergency planning; technical support, including accident and root-cause analysis; and support services such as radiological controls, chemistry, and maintenance [8, 17].

Training and qualification, in particular, permeates essentially all of the changes. Content for control-room operators and other specialties is developed systematically and formally beginning with intensive job-analysis and moving on through feedback of operating experience. It also represents collaboration among training specialists, incumbents, supervision, management, and subject-matter expertise. Of special significance, operator training is augmented greatly by availability for each reactor of a high-fidelity, replica control-room simulator.

The disastrous 1986 accident at the Chernobyl Unit 4 in the former Soviet Union occurred with a reactor system not used elsewhere in the world. However, it did serve to reinforce many of the designs, operations, and management lessons from the earlier TMI-2 accident. It also provided unprecedented insights into severe-accident behavior and served as a catalyst for completion of post-TMI-2 “action plans,” significantly enhanced international cooperation and collaboration in research initiatives and nuclear-power-plant operation and management [8].

Advanced Reactors

The nuclear power industry has been developing and improving reactor technology for more than five decades. Some of the *advanced – next generation* – reactors, which provide hope for revitalizing the nuclear-power option, are starting to be built [2, 3].

Four generations of reactors are commonly distinguished. *GEN-I* reactors were developed in the 1950s–1960s with only a few in the UK still running today. *GEN-II* reactors are typified by the present US fleet and most operated elsewhere worldwide (including those described in the section “[Reactor Systems](#)” with examples in [Table 1](#)). The second-generation nuclear power units have been found to be safe and reliable, but they are being superseded by better designs [2].

GEN-III are the initial “advanced reactors” – the first are in operation in Japan and others are under construction or ready to be ordered. The more cutting-edge units may be termed *GEN-III+*. Safety-based certification of designs is on a national basis. Many are now under licensing review at the US Nuclear Regulatory Commission with some having received Design Certification. In Europe there is some movement toward harmonized requirements for licensing.

GEN-IV reactors are “still on the drawing board.” They will not be operational before 2020 at the earliest [2]. (See the section “[Generation IV Nuclear Reactors](#)”).

GEN-III reactors include, conceptually [2]:

- A standardized design for each type which is intended to expedite licensing, reduce capital cost and reduce construction time, and ultimately compete economically with coal-fired and other generation alternatives
- A simpler and more rugged design, making them easier to operate and less vulnerable to operational upsets
- Higher availability and longer operating life – typically 60 years
- Further reduced risk of serious reactor accidents
- Resistance to serious damage that would allow radiological release from an aircraft impact
- Higher fuel burnup lifetime, reduced outage time, and reduced amount of radioactive waste

The signature difference for *GEN-III* plants is a strong focus on passive, inherent safety features – e.g., gravity, natural convection, or resistance to high temperatures – which require no active controls or operational intervention to avoid accidents in the aftermath of a serious malfunction. This is in contrast to the more traditional (*GEN-II*, e.g., in the section “[Reactor Systems](#)”) reactor safety systems which are “active” in the sense of involving electric or mechanical operation on command. Certainly some current features of *GEN-II* reactors are passive, e.g., negative reactivity feedback and mechanical pressure relief valves which function absent operator action or auxiliary power. However, parallel redundant systems are also necessary. The advanced reactors emphasize inherent or full passive safety depending only on physical phenomena, not on electric power and functioning engineered components. This emphasis is to the extent that some are referred to, probably inappropriately, as *inherently safe*. (See the next section “[Conceptual Advanced Reactors](#)” for three distinctly different examples).

Conceptual Advanced Reactors

Since the early 1990s, advanced reactor concepts using light-water, helium, and liquid-sodium coolants have been under development from the experience base of the current “parent” reactors. An early, conceptual version of each is described here. All include enhanced negative power feedback mechanisms to assist in obtaining shutdown and passive postaccident/post-shutdown cooling mechanisms.

Advanced Light-Water Reactor The conceptual advanced light-water reactors differ from the traditional ones (section “[Light-Water Reactors](#)”) primarily through reduction or elimination of active emergency system components, including large pumps and diesel generators, and implementation of passive features, from reactor control to emergency core cooling functions [18].

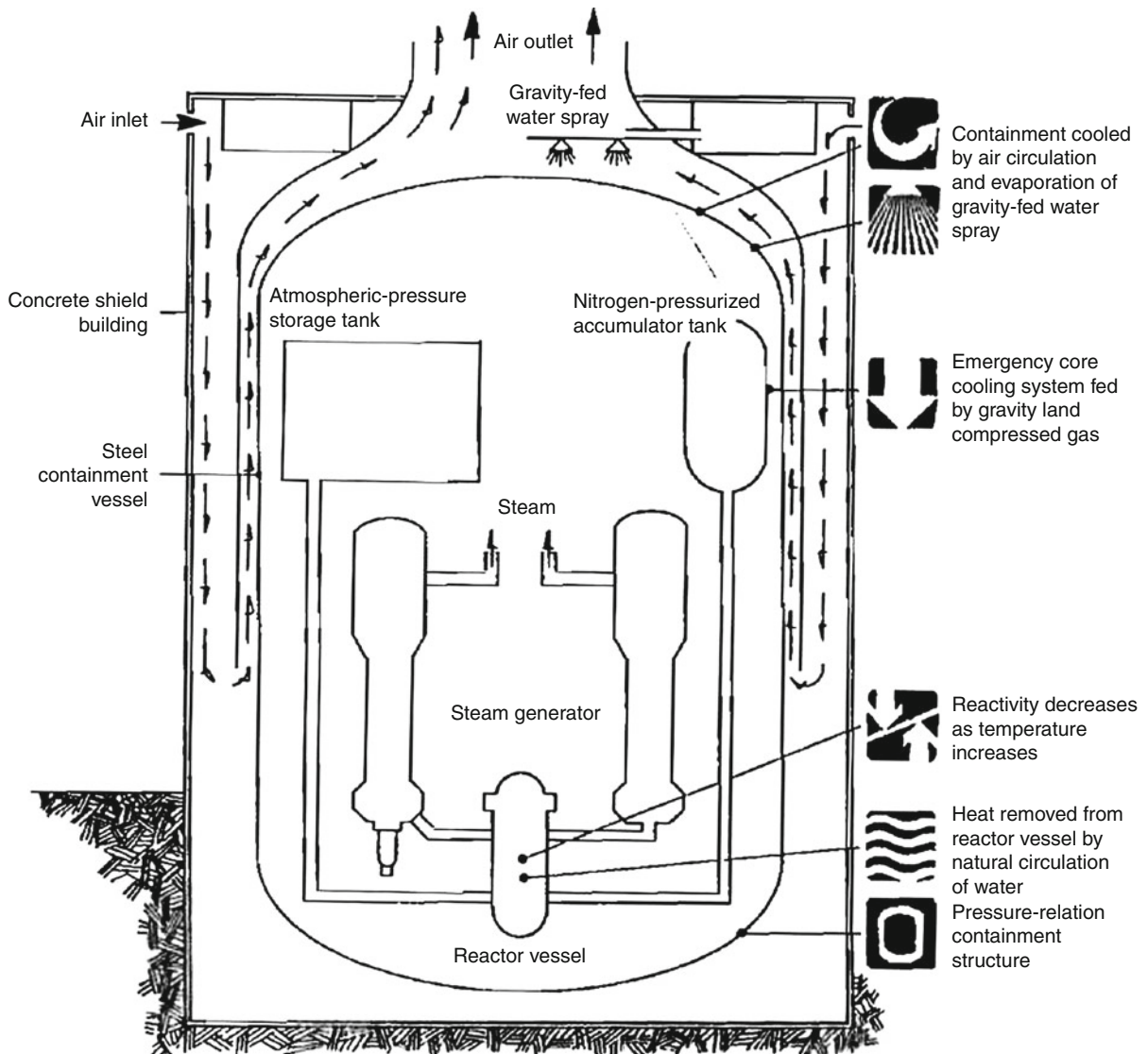
Both *GEN-III* PWR and BWR versions (section “[Light-Water Reactor](#)”) use conventional uranium-oxide fuel assemblies with negative temperature feedback mechanisms. All combat core damage in the aftermath of a loss-of-coolant accident (LOCA) by flooding the core with sufficient water to last for at

least several days – as compared to the current tens of minutes – without further operator action.

An advanced PWR concept (Fig. 25) features passive emergency core cooling systems (ECCS) with water stored in large tanks above the core. During a large LOCA, injection of water into the core from two of the tanks is initiated by pressurized nitrogen even before the reactor coolant system has a chance to depressurize. Subsequent to depressurization, more water from a massive tank inside the containment

structure can flow downward into the core under the influence of gravity alone. Neither of these two ECCS options requires pumps or emergency electric power. If the steam generators are not operable, however, natural circulation of water transfers core thermal energy to a large water storage tank located above the reactor vessel.

Post shutdown, decay heat is ordinarily removed through the steam generators. Heat removal from the steel containment shell is facilitated by



Nuclear Fission Power Plants. Figure 25

Conceptual advanced PWR (Electric Power Research Institute)

a gravity-fed water spray. Natural circulation of air, directed by large baffles, provides passive flow, i.e., without using the large fans and coolers typical of current PWRs.

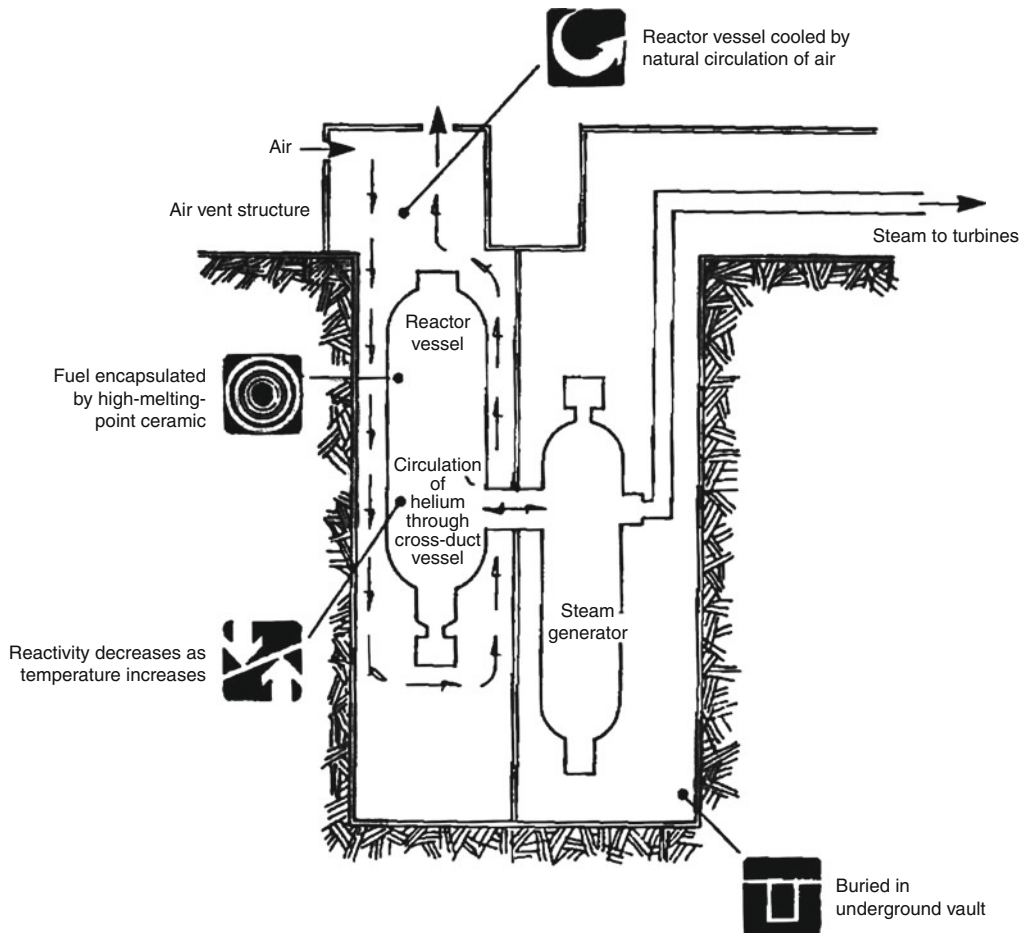
The passive safety systems also simplify overall plant design. Compared with the active safety systems of a current plant, e.g., there are only half as many large pumps and 40% as much in terms of building volume designed to nuclear-grade seismic standards, number of valves, and amount of piping.

Gas-Cooled Reactors The reference HTGR (section “Gas-Cooled Reactors”) combines helium coolant, microsphere fuel, and graphite-moderator blocks to provide uniquely robust safety characteristics. A smaller *modular high-temperature gas-cooled reactor*

(MHTGR) design was built upon these same features. The basic MHTGR (Fig. 26) is a 400–600 MWe module intended to be used in groups of four with a pair of turbine generators.

The graphite and fuel core operates at low power density, has a large inherent heat sink, and is very slow to overheat. Even with depressurization of the primary system, fuel failure would be unlikely. Operator response time, thus, is essentially unlimited.

The helium coolant is circulated through the steel – rather than HTGR’s prestressed concrete – reactor vessel by an electric blower. It operates at high temperature for high steam cycle electric-generation efficiency or, potentially, a variety of industrial processes including coal gasification, chemical manufacturing, petroleum refining, or desalination of sea water.



Nuclear Fission Power Plants. Figure 26

Conceptual advanced reactor MHTGR (Electric Power Research Institute)

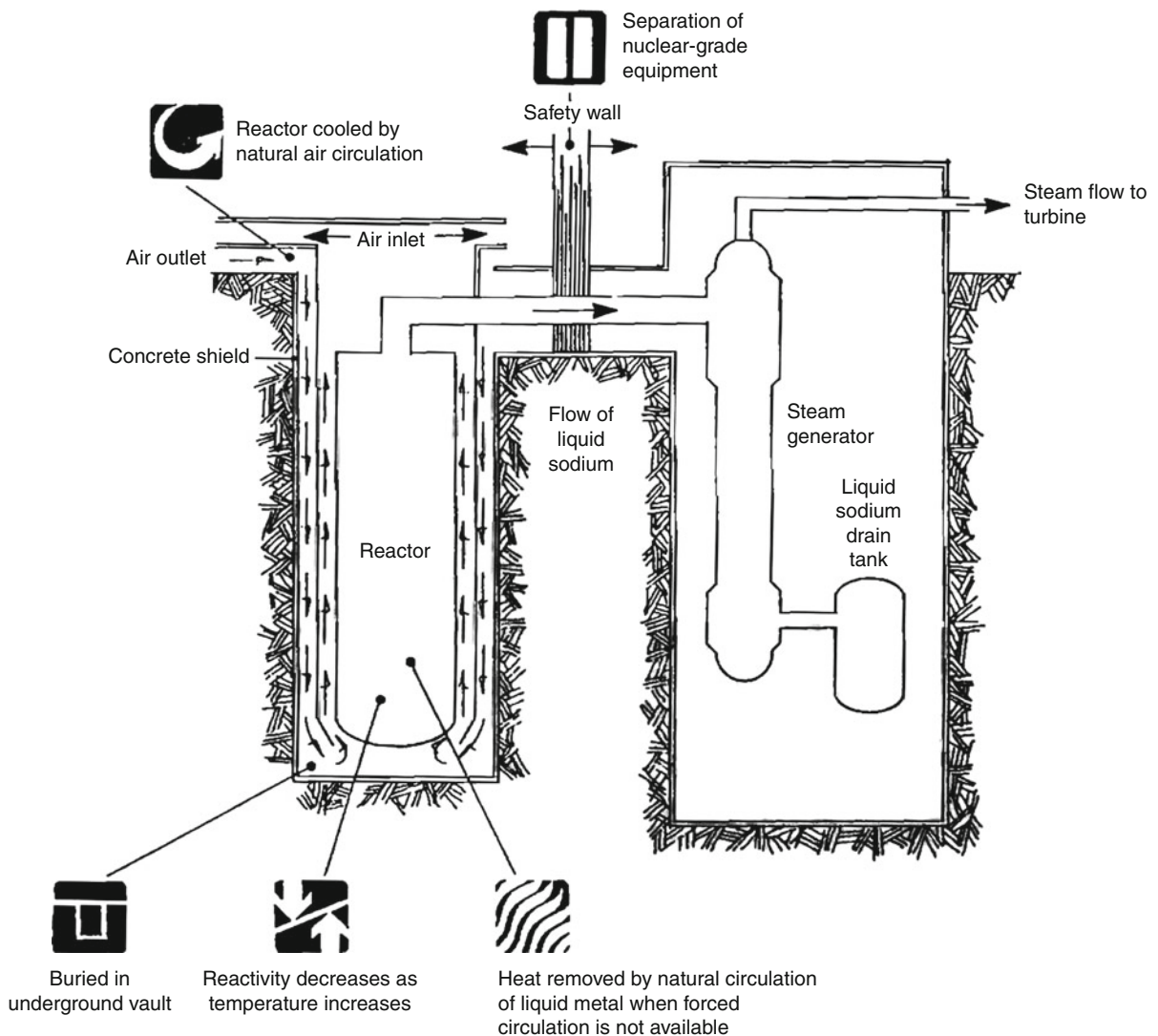
Each reactor vessel and steam generator is enclosed in an underground silo (Fig. 26). Natural air circulation is sufficient to provide passive silo cooling. Even if said circulation were blocked, direct heat loss to the surrounding earth would keep fuel temperatures well below the melting point. Thus, as with the classic HTGR, an LWR-type containment is unnecessary.

Liquid-Metal Reactor Traditionally, liquid-metal reactors have focused on LMFR-like features (section “Light-Water Graphite Reactors”). The power reactor

inherently safe module (PRISM) is a small liquid-metal-cooled module (Fig. 27) which could be used in groups of three connected to a common steam turbine.

Each PRISM module has a pool-type configuration with the liquid-sodium coolant circulated through the core to an intermediate heat exchanger by a pair of pumps. As in an LMFBFR, a secondary sodium loop connects to a steam generator.

If pumps fail, sodium flows through the PRISM core by natural convection. If the secondary flow is



Nuclear Fission Power Plants. Figure 27

Conceptual advanced reactor LMR (Electric Power Research Institute)

interrupted, the primary sodium continues to cool the core by carrying heat to the containment vessel.

Fuel assemblies are made of a pyrometallurgical alloy. Reactor experiments have demonstrated that with loss of liquid-sodium flow, temperature feedback will shut down the neutron chain reaction core and natural circulation of the coolant can remove the decay heat sufficiently to prevent fuel melting.

The PRISM reactor vessel is surrounded by a guard vessel to catch leaking sodium. Both vessels are placed in an underground concrete silo. Air, allowed to circulate freely between the silo wall and guard vessel, can remove core decay heat passively to the outside environment, if necessary. Due to PRISM's passive stability, LWR-type containment is not included in the design.

Generation-III Reactors

The nuclear industry is developing a full product line touted as of clean, safe, secure carbon-free energy sources. Several advanced designs can be ready to meet USA's generating needs by 2015. These fall into the category of "evolutionary," i.e., modifications of existing systems building directly on the experience of operating reactors in the USA, Japan, and Western Europe. These reactors tend to be as large as, or larger than, their predecessors [2, 10].

The industry also is developing more highly advanced nuclear reactors based on new technologies, ready for commercial use by 2020 or later. A revolutionary generation of smaller, passively stable reactors is envisioned [19]. Some could replace older fossil-fired power stations of similar size – and their associated carbon emissions. The infrastructure, cooling water, rail, and transmission facilities already exist at such facilities.

Alternatively, these same small reactors can be a "right size" match for utilities or countries entering the nuclear arena for the first time. They also could be used to generate electricity for remote locations and/or provide process heat for diverse industrial applications. The designs are poised to capitalize on the benefits of modular construction, ease of transportation, and reduced financing. Modules can be added as needed. Units could operate anywhere from 2 to 10 years on an initial fuel load, depending on the design.

Whether the advanced reactor designs are evolutionary or revolutionary, or something in between, almost all involve multi-organizational, multinational collaboration. By the end of 2006, e.g., three major Western-Japanese alliances had formed to dominate much of the world reactor supply market. Subsequently there have been a number of other international collaborative arrangements initiated among reactor vendors and designers. Circa 2010, reactor suppliers in North America, Japan, Europe, Russia, and elsewhere are marketing a dozen new thermal-neutron reactor designs as summarized in Table 3. Others designs, including a number of them based on fast-neutron chain reactions, are, respectively, in advanced planning stages or still in research and development.

Light-Water Reactors Many of the advanced reactors are light-water reactor designs [2]. (See, e.g., Table 3) In the USA, two advanced light-water reactors fall into the category of large "evolutionary" designs, building directly on operating experience in USA, Japan, and Western Europe. Advanced boiling-water reactors (ABWR) derived from a general electric design (section "Boiling-Water Reactors") are in commercial operation and/or under construction in Japan and Taiwan, with others planned in Japan and the USA. It, and the System 80+ advanced PWR, received final design certification from US-NRC in May 1997, noting that they exceeded NRC "safety goals by several orders of magnitude." The ABWR has also been certified as meeting European requirements for advanced reactors. Although the System 80+ itself is not now being promoted for sale, eight System 80 reactors in South Korea incorporate many of its design features and are the basis of the Korean's APR-1400 which has enhanced safety and seismic robustness, and is expected to be in operation by 2013 and is being marketed worldwide.

Another, more innovative US advanced reactor is the smaller AP600 (AP = Advanced Passive) at 600 MWe with highly evolved passive safety features compared to those described in concept in the section "Advanced Light-Water Reactor" (and in concert with Fig. 25). Projected core-damage frequency is nearly 1,000 times less than today's NRC requirements. In 1999, it was the first to gain NRC generic final design certification. Said certifications are valid for 15 years and will allow US utilities to obtain a single NRC

Nuclear Fission Power Plants. Table 3 Advanced thermal-neutron reactors being marketed

Country (developer)	Reactor	Size (MWe)	Design progress	Main features (improved safety in all)
USA-Japan (GE Hitachi, Toshiba)	ABWR	1,300	Commercial operation in Japan since 1996–1997. In USA: NRC certified 1997, FOAKE	<ul style="list-style-type: none"> • Evolutionary design • More efficient, less waste • Simplified construction (48 months) and operation
South Korea (KHNP, derived from Westinghouse)	APR-1400 (PWR) from System80+	1,450	Design certification 2003, first units expected to be operating c. 2013	<ul style="list-style-type: none"> • Evolutionary design • Increased reliability • Simplified construction and operation
USA(Westinghouse)	AP-600 AP-1000 (PWR)	600 1,100	AP-600: NRC certified 1999, FOAKE AP-1000 NRC certification 2005, under construction in China, many more planned there	<ul style="list-style-type: none"> • Simplified construction and operation • 3 years to build • 60-year plant life
USA(GE Hitachi)	ESBWR	1,550	Developed from ABWR, under certification in USA, likely construction there	<ul style="list-style-type: none"> • Evolutionary design • Short construction time
Japan(Utilities, Mitsubishi)	APWR US-APWR EU-APWR	1,530 1,700 1,700	Basic design in progress, planned for Tsuruga US design certification application 2008	<ul style="list-style-type: none"> • Hybrid safety features • Simplified construction and operation
France-Germany (Areva NP)	EPR US-EPR (PWR)	1,600	Future French standard French design approval Being built in Finland, France, and China Under certification in USA	<ul style="list-style-type: none"> • Evolutionary design • High fuel efficiency • Flexible operation
Germany (Areva NP)	SWR-1000 (BWR)	1,200	Under development, pre-certification in USA	<ul style="list-style-type: none"> • Innovative design • High fuel efficiency
Russia (Gidropress)	VVER-1200 (PWR)	1,200	Under construction at Leningrad and Novovoronezh plants	<ul style="list-style-type: none"> • Evolutionary design • High fuel efficiency • 50-year plant life
Canada (AECL)	Enhanced CANDU-6	750	Improved model licensing approval 1997	<ul style="list-style-type: none"> • Evolutionary design • Flexible fuel requirements
Canada (AECL)	ACR	700 1,080	Undergoing certification in Canada	<ul style="list-style-type: none"> • Evolutionary design • Light-water cooling • Low-enriched fuel
China (INET, Chinergy)	HTR-PM	2 × 105 (module)	Demonstration plant due to start building at Shodaowan	<ul style="list-style-type: none"> • Modular plant, low cost • High temperature • High fuel efficiency
USA-Russia et al. (General Atomics – OKBM)	GT-MHR	285 (module)	Under development in Russia by multinational joint venture	<ul style="list-style-type: none"> • Modular plant, low cost • High fuel efficiency • Direct cycle gas turbine

Source: Adapted from World Nuclear Association – <http://www.world-nuclear.org/info/inf08.html>

license to both construct and operate a reactor before construction begins.

The scaled-up AP1000 was the first GEN-III+ type reactor to receive NRC final design certification from the NRC in December 2005. It also is under active consideration for building in Europe and USA. In comparison to current plants for the same size, AP1000 is about one quarter the size with concrete and steel requirements are less by a factor of five. It is designed for unique modular construction. A single unit has 149 structural modules of five kinds, and 198 mechanical modules of four kinds – equipment, piping and valve, commodity, and standard service modules. These comprise one third of all construction and can be built off site in parallel with the on-site construction. China has four AP-1000 units under construction, with many more to follow.

GE Hitachi Nuclear Energy's ESBWR is a GEN-III+ BWR technology that utilizes passive safety features and natural-circulation principles. It is essentially a larger evolution from a predecessor SBWR.

Mitsubishi's large – advanced – APWR was developed in collaboration with four utilities. It is a simpler four-loop design which combines both active and passive cooling systems. The first two are planned for Japan to come on-line from 2016 and will be the basis for the next generation of Japanese PWRs. US-APWR and EU-APWR versions, respectively, are being developed for USA and European markets.

In Europe, several designs are being developed to meet the European Utility Requirements (EUR) of French and German utilities, which have stringent safety criteria. European pressurized-water reactor (EPR) developed from French and German designs has four separate, redundant safety systems rather than passive safety. The first EPR unit is being built in Finland, with two other slated for France, and another for China. US-EPR – now known as the Evolutionary PWR (EPR) – was submitted for US design certification.

Another evolutionary design, the SWR 1000, is a Siemens BWR now known as Kerena. It is simpler overall and has many passive safety features.

In Russia, several advanced PWR designs with passive safety features have been developed. Late-model VVER-1000 units (see the section "[Pressurized-Water Reactor](#)") – one considered GEN-III with enhanced

safety – are being built in India and China with two more planned for Bulgaria. A third-generation standardized VVER-1200 reactor is an evolutionary development from the VVER-1000 with longer life, greater power, and greater efficiency. They have enhanced safety including that related to earthquakes and aircraft impact with some passive safety features, double containment, and low core-damage frequency. A Europe-tailored reactor design, MIR-1200 – Modernized International Reactor – has some Czech involvement.

Russia also has a number of other novel small reactor designs [2, 19]. Several VBER PWR variations developed from naval power plants may be floating or land-based, modular construction, and/or provide cogeneration. VK-300 – with fuel elements similar to the VVER – have been developed specifically for cogeneration of both power and district heating or heat for desalination. Cooling (by convection) and all safety systems are passive. It has been announced that six would be built in Russia's far east.

Heavy-Water Reactors A pair of GEN-III HWRs are under development based directly on CANDU reactors. Each is to be configured for flexible fuel requirements including standard natural uranium, slightly enriched LWR-like uranium, and mixed-oxide (U and Pu) fuel and from a variety of potential sources, e.g., PWR spent-fuel or even military down-sizing.

Such fuel innovations, combined with the experiences of South Korea and China, suggest that an Enhanced CANDU-6 (EC6) could have a 4.5-year construction and 60-year plant life. This design is under consideration for new build in Canada.

The Advanced CANDU Reactor (ACR) is a more innovative concept using a low-pressure heavy-water moderator with PWR-like light-water cooling for higher thermal efficiency. The resulting compact core would reduce capital cost – an ACR-700 is physically much smaller, simpler, and more efficient as well as 40% cheaper. A larger ACR-1000 – the current focus of attention – has more fuel channels; is designed to run on low-enriched uranium ($\sim 1.5\text{--}2.0\text{ wt}\% \text{ }^{235}\text{U}$) with high burnup, MOX fuel, thorium- ^{233}U fuels; and extends the fuel life by about three times reducing high-level waste volumes accordingly. Other features include small negative reactivity feedback, other

passive safety features, two independent and fast shut-down systems; and prefabricated modules to cut construction time to 3.5 years. The CANDU X or SCWR – a GEN-IV ACR variant with supercritical light-water coolant – would operate at very high pressure and temperature, and, thus, high thermal efficiency. (See the section “[Generation IV Nuclear Reactors](#)”.)

[Table 3](#) summarizes key characteristics for the advance HWRs and other advanced reactor types.

For the longer term, India is developing the Advanced Heavy Water Reactor (AHWR) to utilize its plentiful reserves of thorium. The AHWR is a 300-MWe reactor moderated by heavy water in a calandria vessel with about 450 *vertical* pressure tubes. The coolant is boiling light water circulated by convection. A “gravity-driven water pool” near the top of the reactor building with 7,000 cubic meters of water is proposed as a novel heat sink. A unique fuel paradigm calls initially for fuel assemblies each consisting of a mixture of oxide pins of Th-²³³U and Pu-Th, respectively, with resulting very high net fuel utilization. Once fully operational, the fuel pins would be arranged in three concentric rings each “zoned” with a specific equivalent fissile content based on adjusting Th, ²³³U, and Pu fractions. The content of ²³³U – with inherent ²³²U and its unique high-gamma-active daughter products – is considered to confer a substantial proliferation resistance (i.e., high-capacity fabrication of ²³³U fuel must be done remotely, a feature not necessary for either ²³⁵U or plutonium). (See also the chapter on [► Modern Nuclear Fuel Cycles](#).)

In 2009, an export version named AHWR-LEU was announced. The major difference is use of low-enriched uranium (LEU) – in lieu of Pu – plus thorium as a fuel. Plutonium production will differ from that in light-water reactors – the fissile proportion will be less and the ²³⁸Pu (non-fissile with high radioactivity and associated heat) portion three times as high – giving inherent proliferation resistance. The net result is described as a reactor “manageable with modest industrial infrastructure within the reach of developing countries.”

High-Temperature Gas-Cooled Reactors The advanced high-temperature gas-cooled reactors (HTGR) use helium as a coolant at up to 950°C, which for electric generation either makes steam conventionally or directly drives a gas turbine for electricity

and a compressor to return the gas to the reactor core. They also are especially well-suited for providing process heat for industrial applications, including hydrogen production and could be used in the development of tar sands, oil shale, and coal-to-liquid applications, reducing the life cycle carbon footprint of all these activities.

The fuel is in the conventional HTGR form of microsphere particles arranged in blocks as hexagonal “prisms” of graphite, or in billiard ball-sized “pebbles” of graphite encased in silicon carbide ([Fig. 15](#)) each providing containment for fission products which is stable to 1,600°C or more. Other features are similar to those in [Fig. 26](#).

China’s HTR-PM is to have two reactor modules of 105 MWe each and use a steam cycle to drive a single steam cycle turbine at about 40% thermal efficiency. It is expected to pave the way for an 18-unit full-scale power plant also using the steam cycle.

A larger US design, the Gas Turbine–Modular Helium Reactor (GT-MHR), is planned as modules of 285 MWe each directly driving a gas turbine (Brayton cycle) at 48% thermal efficiency. The cylindrical core consists of 102 prism blocks ([Fig. 15b](#)) with channels for helium and control rods. It is being developed by General Atomics in partnership with Russia, and supported by Japan. Initially it was to be used to burn pure ex-weapons plutonium in Russia. The preliminary design stage was completed in 2001, but the program has stalled since. Areva’s Antares also is based on the GT-MHR.

[Table 3](#) summarizes key characteristics for the first two of the advanced HTGRs and for other advanced reactor types.

South Africa’s Pebble Bed Modular Reactor (PBMR) draws on German THTR (section “[Gas-Cooled Reactor](#)” and [Fig. 14b](#)) expertise. Production units would be 165 MWe, ultimately using a direct-(Brayton) cycle gas-turbine generator with thermal efficiency of about 41%. Power would be adjusted by changing the pressure in the system.

Up to 450,000 fuel pebbles ([Fig. 15a and c](#)) recycle through the reactor continuously (about six times each) until they are expended. The reactor will use about 13 fuel loads in a 40-year lifetime. The on-line refueling can facilitate operational cycles with as long as 6 years between shutdowns.

Fast-Neutron Reactors The classic fast-neutron reactor has been the fast breeder reactor (FBR). Current interest centers on their ability to provide unique nuclear fuel-cycle services, such as breeding new fuel, consuming recycled nuclear waste as fuel, and supporting government-sponsored nonproliferation efforts by consuming material from former nuclear weapons [10].

About 20 liquid-metal fast breed reactors (LMFBR), mainly research units, have accumulated some significant operating experience. Several countries have research and development programs for improving their current designs, but generally less as GEN-III efforts and more as a bridge to new GEN-IV reactors [2, 20].

The consummate LMFBR, France's SuperPhenix (section "[Light-Water Graphite Reactors](#)"), was optimized to run on plutonium fuel and to "breed" even more plutonium fuel with a depleted uranium blanket around the core. The LMFBRs, however, are expensive to build and could only be justified economically if uranium prices were to rise dramatically above market price circa 2010. For this reason, research work almost ceased for some years. Closure of the SuperPhenix after very little operation over 13 years especially set back developments.

Primarily to address nonproliferation concerns, there apparently are no plans to build any more plutonium-fueled fast reactors with blanket assemblies. Fast-reactor concepts slated for the GEN-IV program simply have a core where both plutonium production and consumption occur [20].

In Russia, the BN-600 – a fast breeder reactor with a conventional core-blanket configuration – has supplied electricity to the grid since 1981 (and has had the best operating and production record of all of the nation's nuclear power units). Its evolutionary BN-800 counterpart originally was fueled with uranium-oxide fuel, but reconfigured to burn the plutonium from military stockpiles. Two BN-800 units have been sold to China with construction slated to start in 2012. Further development of the BN-series is slated to be with an integrated core to meet GEN-IV nonproliferation goals. Russia also has interest in developing lead- and lead-bismuth fast-neutron-reactor designs, having used these coolants in submarines.

Japan is continuing with development of the FBR. The Japan Standard Fast Reactor (JSFR) is

a breeder-like concept, but having a breeding ratio less than unity.

Meanwhile in the USA, the modular liquid-metal-cooled inherently safe reactor – PRISM (See the section "[Liquid-Metal Reactors](#)" and [Fig. 27](#)) – is continuing to receive attention. It is a compact modular pool-type reactor with passive cooling for decay heat removal. PRISM is considered a potential GEN-IV solution to closing the fuel cycle in the USA.

Korea's KALIMER (Korea Advanced LIquid METal Reactor) also is pool-type sodium-cooled fast reactor. Having evolved from a smaller unit, it has a core designed to transmute transuranic species and no breeding blanket. Future development of KALIMER for GEN-IV is envisaged.

Research continues in India including on the unique track of using thorium – which it has in abundance – in a breeding scheme with fissile ^{233}U . The Kalpakkam reactor is fueled with uranium-plutonium carbide and has a thorium blanket to breed the ^{233}U .

Generation IV Nuclear Reactors

The GEN-IV International Forum (GIF) was initiated in 2000 as an international collective committed to joint development of the next generation of nuclear technology. Currently six nuclear reactor technologies are under development for deployment between 2020 and 2030 [21, 22].

All six technologies represent advances in sustainability, economics, safety, reliability, and proliferation resistance. Most employ a closed fuel cycle to maximize the resource base and minimize high-level wastes that will need to be sent to a repository.

Collectively, the reactors include thermal-, epithermal-, or fast-neutron spectra. Coolants are light water, helium, sodium, lead-bismuth, and fluoride salt. Temperatures range from 510°C to 1,000°C – compared with less than 330°C for today's LWRs. At the higher temperatures, thermochemical hydrogen production is envisioned. Power ratings range all the way from 20 to 1,500 MWe (or equivalent thermal).

At least four of the systems already have significant operating experience in most respects of their design. This provides a good basis for further research and development and prospects for commercial operation well before 2030.

Table 4 provides a comparison. (See also the following section “[Future Directions](#)”.) Although six technologies were chosen, development on one of them has gone in two different directions resulting in seven being listed in the table. Summary descriptions of each are provided in the remainder of this section.

Supercritical water-cooled reactors (SWCR) use water coolant at extremely high temperatures and pressures – thermodynamically in the “supercritical” fluid regime. Such fluids are those above the thermodynamic critical point, defined as the highest temperature and pressure at which gas and liquid phases can coexist in equilibrium. They have properties between those of gas and liquid. For water the critical point is at 374°C and 22 MPa, giving it a “steam” density one third that of the liquid so that it can drive a turbine in a similar way to normal steam. A fleet of over 400 supercritical coal-fired plants operating worldwide provides a solid experience base for reactor application [21].

Supercritical water obviates the need for a secondary steam system and drives a turbine generator directly at 44% (and potentially up to 50%) efficiency. Designs call for a high power level and include BWR-like pressure-vessel and CANDU-variant pressure-tube options. Each will have passive safety features with fewer components (e.g., none related to steam separator/driers) and relatively low capital cost. Neutron-spectrum and fuel-cycle options combine in an interesting manner to provide the potential for:

1. Thermal reactors with enriched uranium-oxide fuel and an open fuel cycle
2. Fast reactors with full actinide-element recycle and conventional reprocessing

Very high-temperature gas reactors (VHTR) are thermal-neutron systems with graphite moderator and helium coolant. Fuel can be (1) stacks of prism blocks or (2) a pebble bed (section “[Gas-Cooled Reactors](#)” and Fig. 15). The “very high” temperatures result in high thermodynamic efficiency which can be applied for thermochemical hydrogen production with electricity cogeneration or direct high-efficiency driving of a gas turbine (Brayton cycle). Essentially complete passive safety, low operation and maintenance costs, and modular construction are envisioned based on the HTGR (section “[Gas-Cooled Reactors](#)”) and PBR

(section “[High-Temperature Gas-Cooled Reactors](#)”) heritage.

Gas-cooled fast reactors (GFR) are the fast-neutron counterparts of the VHTR – absent graphite moderator – described above. The design includes a thick steel reactor pressure vessel and helium coolant. It is well suited for conventional steam-electric generation, helium direct-drive gas turbine (Brayton cycle), and thermochemical hydrogen production or other process heat. With the fast-neutron spectrum, fuel of Pu, ^{238}U , and Th can support a blanket-less breeding paradigm.

Sodium-cooled fast reactors (SFR), using liquid-sodium coolant, are progeny of the LMFBR (section “[Light-Water Graphite Reactors](#)”). Configurations could be loop-type or pool-type. Passive safety features are augmented by the coolant’s low pressure and inherent natural-convection tendency. An initial core plus blanket configuration is likely to be displaced by new core-only breeding designs with depleted uranium integral to the fuel matrix. High-temperature operation is consistent with electricity generation or a supercritical CO₂ turbine. Fuel-related variants include: (1) U-Pu metal fuel with actinide retention and on-site pyrometallurgical processing and (2) conventional mixed Pu-U “mixed-oxide” (MOX) fuel and advanced aqueous reprocessing elsewhere in central facilities.

Lead-cooled fast reactors (LFR) have similarity to the SCFR in terms of fast-neutron spectrum, as well as liquid-metal coolant – here Pb or Pb-Bi eutectic – with low pressure and natural convection. Russia has had multi-decade experience with this technology in its nuclear submarine program. Design power levels range from relatively small for modular units to large for single plants. A novel concept is for a factory-built “battery” unit of 20–180 MWe with 15–20 year life for small grids or developing countries. Operating temperatures are slated to be sufficient for thermochemical hydrogen production. Fuel would be Pu and depleted uranium – with an option to introduce thorium – in metal or nitride form.

Molten-salt reactors (MSR) – the most novel of the technologies – use uranium dissolved in sodium (or zirconium) fluoride salt which serves as both the fuel and the coolant. When the salt circulates through graphite core channels, the neutrons are moderated to epithermal (i.e., average above, or faster than, thermal)

Nuclear Fission Power Plants. Table 4 Generation IV nuclear reactors

	Neutron spectrum (fast/thermal)	Coolant	Temperature (C)	Pressure ^a	Fuel	Fuel cycle	Size(s) (MWe)	Uses
Supercritical water-cooled reactors (SCWR)	Thermal or fast	Water	510–625	Very high	UO ₂	Open (thermal)	300–700	Electricity
						Closed (fast)	1,000–1,500	
Very high temperature gas reactors (VHTR)	Thermal	Helium	900–1,000	High	UO ₂ prism or pebbles	Open	250–300	Hydrogen and electricity
Gas-cooled fast reactors (GFR)								
Sodium-cooled fast reactors (SFR)	Fast	Sodium	550	Low	U-238 and MOX	Closed	30–150 300–1,500 1,000–2,000	Electricity
Lead-cooled fast reactors (LFR)	Fast	Lead or Pb-Bi	480–800	Low	U-238 ^b	Closed, regional	20–180 ^c 300–1,200 600–1,000	Electricity and hydrogen
Molten salt fast reactors (MSFR)	Fast	Fluoride salts	700–800	Low	UF in salt	Closed	1,000	Electricity and hydrogen
Molten salt advanced high-temperature reactors (AHTR)	Thermal	Fluoride salts	750–1,000		UO ₂ particles in prism	Open	1,000–1,500	Hydrogen

Source: Adapted from World Nuclear Association – <http://www.world-nuclear.org/info/inf77.html>^aHigh 7–15 MPa^bWith some U-235 or Pu-239^c“Battery” model with long cassette core life (15–20 years) or replaceable reactor module

energies and can support a critical chain reaction. Outside of the core and absent proximity to moderator, the salt is very highly subcritical. Similar to the metal-cooled reactors, the molten salt has the inherent safety feature of operating at very low pressure. An appropriately designed secondary loop can provide for electricity generation or thermochemical hydrogen production.

The MSR fuel requires no fabrication per se. Fissile inventory can be kept to a minimal level as the plutonium, depleted uranium, and/or thorium fuel constituents can be inserted at any time as needed. Actinides from outside sources, e.g., to be transmuted by neutron irradiation to less hazardous form, can also be added. Simultaneously, on-line and continuous removal of fission products and recycle of self-generated actinides can be implemented.

There are two separate baseline MSR concepts. One is the (Molten Salt) Advanced High-Temperature Reactor (AHTR) which has a graphite core similar to that of the VHTR (and HTGR Fig. 15b). Use of a neutron-moderating, molten-salt coolant, instead of helium, supports a much higher power density. The other version is the Molten Salt Fast Neutron Reactor (MSFR) with a graphite-free “core tank with no internals” and molten-salt compositions which provide less moderation.

Future Directions

The future of nuclear power as a world energy source is a complex part of an overall energy picture, a few aspects of which were summarized in the section “[Definition of the Subject](#).” Two recent studies – sponsored, respectively, by the Massachusetts Institute of Technology (MIT) and the Electric Power Research Institute (EPRI) – shed light on both future prospects and problems.

The Future of Nuclear Power

The MIT study, *The Future of Nuclear Power* (2003 report and 2009 update) – in its own words – was “the most comprehensive, interdisciplinary study ever conducted on the future of nuclear energy,” and concluded that “The nuclear option should be retained precisely because it is an important carbon-free source of power” [4].

The study examines a growth scenario – characterized as “not a prediction, but rather a study case in

which nuclear power would make a significant contribution to reducing CO₂ emissions” – with the current deployment of 360 GWe of nuclear capacity worldwide expanded to 1,000 GWe (1 Terawatt) of “carbon-free” power by mid-century, keeping nuclear’s fraction of the electricity market about constant. Correspondingly, the USA would triple its nuclear deployment from about 100 GWe to 300 GWe. (See also the subsequent section “[New Nuclear Installations](#)”).

The credibility of such a scenario – and nuclear energy as an option – will be largely determined in the forthcoming decade by the extent to which significant progress is made on four unresolved problems:

1. High relative construction costs
2. Perceived adverse safety, environmental, and health effects
3. Potential security risks stemming from proliferation
4. Unresolved challenges in long-term management of nuclear wastes

The up-front costs associated with nuclear power are higher than those for fossil-fueled plants. However, the study was optimistic about ways to mitigate the costs, and noted that, over time, the societal and environmental price of carbon emissions could dramatically improve the competitiveness of nuclear power. The goal of 1 terawatt of “carbon-free” nuclear power by 2050 is certainly challenging, requiring deployment of about 2,000 MW a month with an associated capital investment of \$2 trillion over several decades.

One key economic incentive has been a limited production tax credit – a provision in the 2005 Energy Policy Act – to “first mover” private sector investors who successfully build new nuclear plants. This tax credit is extendable to other carbon-free electricity technologies and is not paid unless the plant operates. The study also suggested the industry-wide cost-reduction options of placing near-term emphasis on the once-through fuel cycle and delaying expensive development projects pending progress on near-term deployments.

Public acceptance turns on the perception of safety, environmental, health effects. Thus, performance, cost, and environmental acceptability of the technology must be demonstrated to the public, political leaders, and investors. Having the government develop

enhanced capabilities to analyze life-cycle health and safety impacts of fuel-cycle facilities is also recommended. It also is essential that NRC regulations are developed and enforced diligently.

Security risks can be addressed by encouraging countries to forego the proliferation-risky nuclear technologies – enrichment and reprocessing – by offering “a preferred position to receive nuclear fuel and waste management services from nations that operate the entire fuel cycle.” Near-term emphasis on the once-through fuel cycle also supports this scenario.

Long-term management of nuclear wastes is a huge problem, the solution for which is key to public acceptance. Key to the resolution of this task is the capability of the government to start moving spent fuel from reactor sites and to develop long-term waste management research and development programs. (See also the following section “[New Nuclear Installations](#)” with respect to spent-fuel management.)

According to a 2009 MIT update report [4], the initial 2003 study in the years following its publication was credited with a significant impact on the public debate both in the USA and abroad, including influence on the US Department of Energy’s (DOE) nuclear energy research and development (R&D) program. Unfortunately, in this same time frame, the challenges to greater nuclear-power deployment were found to have remained largely the same – especially with no new US plants under construction and insufficient progress on waste management – leading to: “The sober warning...that if more is not done, nuclear power will diminish as a practical and timely option for deployment at a scale that would constitute a material contribution to climate change risk mitigation.”

New Nuclear Installations

An EPRI summer seminar in 2007 entitled “Electricity Solutions for a Carbon-Constrained Future” addressed “New Nuclear Installations” [6]. Whereas the MIT study [4] focused primarily on policy, the EPRI nuclear effort provides a valuable counterpoint on implementation.

At the time, nuclear power accounted for 73% of the emission-free generation in the USA and was described as the only technologically mature,

non-emitting source of power that positioned to deliver large-scale CO₂ reduction in the upcoming decades. Employing a study target of 64 GW of new nuclear by 2030 – considered ambitious but achievable – key challenges were identified so as to:

1. Keep the current nuclear power plants running safely and reliably for 60–80 years.
2. Build out the next generation of plants starting around 2015.

The existing fleet of light-water reactor (LWR) technology – generating approximately 20% of the nation’s electricity and operating at an average capacity of 90% – establishes a platform of confidence for the nation to proceed with further life extension of existing plants and to considerably expand the fleet using advanced LWR designs.

3. Achieve consensus on a long-term strategy for spent fuel.

Almost all of the US operating plants have had, or are in the process of having, their operating licenses extended from the initial 40–60 years. It is a reasonable assumption that all units will be granted a 20-year life extension by about 2016.

Even while these life extensions are being finalized, the technical basis must be laid for an additional 20-year extension – from 60 to 80 years – by confirming that “...with sufficient maintenance, refurbishment, and upgrades, today’s plants could [continue to] operate quite safely ...” “Ultimately, extending the life of our current fleet an additional 20 years will be a business decision, which means that both continued high safety performance and continued economic competitiveness must be addressed.” Key among a wide variety of important milestones are assessing the ability of passive components to continue safe performance, upgrading instrumentation and controls to modern digital technology, and developing of higher-performance fuels to extend the time between refueling outages and reduce spent-fuel volumes.

Design development and preclicensing have produced advanced light-water reactor (ALWR) designs (section “[Advanced Reactors](#)”) that are approaching the “essentially complete design” status which will enable new plant orders to be based on detailed cost and schedule estimates. ALWRs are already in operation today or under construction around the world.

At least fifteen US companies stated their intent to file for NRC combined construction and operating licenses. Of a total of 30 reactors under consideration, most were slated for sites that have existing units and were developed with such expansion in mind.

The first commercial operation of ALWRs in the USA could begin shortly after 2015. "The first new plants out of the box must be done very, very well. . . . They must be executed thoughtfully, deliberately, and with the highest level of skill." The challenge is daunting. Then, the industry must sustain a much higher build rate out to 2050 and beyond. The research, development, and demonstration (RD&D) – with particular emphasis on the latter – focus for new advanced reactors should/must include:

- Completing, in the short term, engineering work necessary for detailed cost and schedule estimates for plant construction, and resolution of any remaining regulatory issues
- Beginning now to lay the foundation for the high build rates, including bringing capital costs and construction times down, addressing shortfalls in both physical and workforce infrastructure, and dealing with a plethora of other issues
- Helping the US federal government advance the capabilities of the high-temperature gas-cooled reactor (HTGR) from construction of an operational prototype to enabling commercially available units, e.g., to take advantage of the ability to produce electricity, provide process heat for industrial applications, and – especially – cost-effectively generate hydrogen by emission-free methods

Finally, sustained expansion of nuclear generation ultimately requires the resolution of spent-fuel management concerns. Although on-site interim storage in concrete silos has been effective and safe for over two decades, imperatives of economics, security, and sustainability will require establishing an integrated fuel management system for the longer term, i.e., including centralized interim storage, long-term geologic storage, and eventually, a closed fuel cycle. A well-thought-out, deliberate consensus strategy on nuclear fuel storage is needed. Such a consensus will allow for advanced reprocessing and separation technologies, reconditioning, fuel manufacturing facilities, and effective application of GEN-IV fast-reactor technologies.

Notes

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Bibliography

Primary Literature

1. Rhodes R (1986) *The making of the atomic bomb*. Simon and Schuster, New York
2. Advanced nuclear power reactors – World Nuclear Association – <http://www.world-nuclear.org/info/inf08.html>. Accessed 6 July 2011
3. Knief RA (1997) Nuclear steam supply systems. In: Elliott T et al (eds) *Standard handbook of power plant engineering*, 2nd edn. McGraw-Hill Book, New York
4. Ansolabehere S et al (2003) The future of nuclear power – an interdisciplinary MIT study. Massachusetts Institute of Technology. <http://web.mit.edu/nuclearpower/>. [Summarized in Deutch JM, Moniz EJ (2006) *The nuclear option*. Sci. Am, September 2006.] Also "Update of the MIT 2003 future of nuclear power study," Massachusetts Institute of Technology, 2009 – <http://web.mit.edu/nuclearpower/pdf/nuclearpower-update2009.pdf>
5. Lake JA, Bennett RG, Kotek JF (2009) Next generation nuclear power. *Sci Am* 286(1):72–79
6. New nuclear installations (from EPRI's 2007 Summer seminar: electricity solutions for a carbon-constrained future), EPRI J (Fall 2007):26–28
7. 2010 World nuclear industry handbook, from the publishers of Nuclear Engineering International magazine, www.neimagazine.com
8. Knief RA (1992) *Nuclear engineering – theory and technology of commercial nuclear power*, 2nd edn. Taylor & Francis, Washington, DC (Reprinted by American Nuclear Society, 2008; 3rd edition in preparation)
9. Sesonske A (1973) *Nuclear power plant design analysis*. U.S. Atomic Energy Commission, TID-26241, Washington, DC
10. Nuclear Energy Institute – <http://www.nei.org/keyissues/newnuclearplants/> – <http://www.nei.org>. Accessed 6 July 2011
11. World Nuclear Association – <http://www.world-nuclear.org/>. Accessed 6 July 2011
12. UNM (University of New Mexico) CEL (Centennial Engineering Library) Nuclear Engineering Wallcharts – <http://econtent.unm.edu/cdm4/browse.php?CISOROOT=/nuceng> – Nuclear Engineering International – <http://www.neimagazine.com/story.asp?storyCode=2055665>. Accessed 6 July 2011
13. Rahn FJ, Adamantiades AG, Kenton JE, Braun C (1984) *A guide to nuclear power technology: a resource for decision making*. Wiley-Interscience, New York

14. U.S. Nuclear Regulatory Commission (1975) Reactor safety study: an assessment of risks in U.S. commercial nuclear power plants. U.S. Nuclear Regulatory Commission, WASH-1400 (NUREG-74/014), Washington, DC
15. U.S. Nuclear Regulatory Commission (1989) Severe accident risks: an assessment for five U.S. nuclear plants. U.S. Nuclear Regulatory Commission, NUREG-1150, Washington, DC
16. Nuclear Engineering International magazine – <http://www.neimagazine.com/>. Accessed 6 July 2011
17. Walker SJ (2004) Three Mile Island – a nuclear crisis in perspective. University of California Press, Berkeley
18. Catron J (1989) New interest in passive reactor designs. EPRI J 14(3):4–13
19. Small Nuclear Power Reactors – World Nuclear Association – <http://www.world-nuclear.org/info/inf33.html>. Accessed 6 July 2011
20. Fast Neutron Reactors – World Nuclear Association – <http://www.world-nuclear.org/info/inf98.html>. Accessed 6 July 2011
21. Generation IV Nuclear Reactors – World Nuclear Association – <http://www.world-nuclear.org/info/inf77.html>. Accessed 6 July 2011
22. Generation IV Nuclear Energy Systems – U.S. Department of Energy – <http://www.ne.doe.gov/genIV/neGenIV1.html>. Accessed 6 July 2011
23. Seminov BA (1983) Nuclear power in the Soviet Union. IAEA Bull 25(2):47–59
- Marcus GH (2010) Nuclear firsts: milestones on the road to nuclear power development. American Nuclear Society, La Grange Park
- Marshall W (ed) (1983) Nuclear power technology – volumes 1, 2, and 3. Clarendon Press, Oxford
- McIntyre MC (1975) Natural uranium heavy-water reactors. Sci Am 233:17–27
- MIT: the future of nuclear power. <http://web.mit.edu/nuclearpower/>. <http://web.mit.edu/nuclearpower/pdf/nuclearpower-update2009.pdf>. Accessed 6 July 2011
- Murray RL (2008) Nuclear energy, 6th edn. Butterworth-Heinemann, Oxford
- Nero AV Jr (1979) A guidebook to nuclear reactors. University of California Press, Berkeley
- Nuclear Engineering International Special Publications (1989) French PWR technology. Nuclear Engineering International Special Publications, Sutton
- Organisation for Economic Co-Operation and Development (OECD) – Nuclear Energy Agency (NEA). <http://www.nea.fr/>
- Rogovin M, Frampton GT Jr (1980) Three Mile Island: a report to the Commissioners and to the public. U. S Nuclear Regulatory Commission, Washington, DC
- Todreas NE, Kazimi MS (1990) Nuclear systems (vol 1: Thermal hydraulic fundamentals and vol 2: Elements of thermal hydraulic design). Taylor & Francis/Hemisphere, New York
- U.S. Department of Energy (1987) Overall plant design descriptions VVER water-cooled, water-moderated energy reactor. DOE/NE0084, Rev. 1, October
- U.S. Department of Energy (DOE) – Office of Nuclear Energy (NE). www.ne.doe.gov. Accessed 6 July 2011
- U.S. Nuclear Regulatory Commission (1975) Reactor safety study: an assessment of risks in U.S. commercial nuclear power plants, WASH-1400 (NUREG-74/014). U.S. Nuclear Regulatory Commission, Washington, DC
- U.S. Nuclear Regulatory Commission (1987) Report on the accident at the Chernobyl nuclear power station, NUREG-1250, rev.1. U.S. Nuclear Regulatory Commission Washington, DC
- U.S. Nuclear Regulatory Commission (1989) Severe accident risks: an assessment for five U.S. nuclear plants, NUREG-1150. U.S. Nuclear Regulatory Commission, Washington, DC
- U.S. Nuclear Regulatory Commission (NRC). www.nrc.gov. Accessed 6 July 2011
- Vendryes GA (1977) Superphénix: a full-scale breeder reactor. Sci Am 236:26–35
- World Nuclear Association (WNA). <http://www.world-nuclear.org/info/reactors.html>; Advanced Nuclear Power Reactors. <http://www.world-nuclear.org/info/inf08.html>; Generation IV Nuclear Reactors. <http://www.world-nuclear.org/info/inf77.html>; Fast Neutron Reactors. <http://www.world-nuclear.org/info/inf98.html>; Small Nuclear Power Reactors. <http://www.world-nuclear.org/info/inf33.html>. Accessed 6 July 2011

Books and Reviews

- Agnew HM (1981) Gas-cooled nuclear power reactors. Sci Am 244:55–63
- American Nuclear Society (+ Nuclear News). www.ans.org
- Benedict M, Pigford TH, Levi HW (1981) Nuclear chemical engineering, 2nd edn. McGraw-Hill, New York
- Cochran RG, Tsoulfanidis N (1990) The nuclear fuel cycle – analysis and management. American Nuclear Society, La Grange Park
- Douglas J (1994) Reopening the nuclear option. EPRI J 19(8):6–17
- Electric Power Research Institute (EPRI). <http://my.epri.com>. Accessed 6 July 2011
- Fishetti MA (1987) Inherently safe reactors: they'd work if we'd let them. IEEE Spect 24(4):28–33
- Framatome (1992) P4: the 1300 MWe PWR Series. Paris, France, April 1992
- Glasstone S, Sesonske A (1994) Nuclear reactor engineering (volume 1: reactor design basics and volume 2: reactor systems engineering), 4th edn. Chapman & Hall, New York
- Golay MW, Todreas NE (1990) Advanced light water reactors. Sci Am 262(4):82–89
- Hore-Lacy I (2009) Nuclear energy in the 21st century, 2nd edn. World Nuclear University Press, London
- International Atomic Energy Agency (IAEA). <http://www.iaea.org>
- Lewis HW (1980) The safety of fission reactors. Sci Am 242:53–65
- Lish KC (1972) Nuclear power plant systems and equipment. Industrial Press, New York

Nuclear Fuel, Reprocessing of

MICHAEL F. SIMPSON, JACK D. LAW

Fuel Cycle Science and Technology Division, Idaho
National Laboratory, Idaho Falls, ID, USA

Article Outline

Glossary
Objective
Introduction
Reprocessing Technology
Future Directions
Bibliography

Glossary

Actinides All elements including and beyond actinium ($Z > 89$) in the periodic table. In spent fuel, the major actinides of interest are uranium, plutonium, neptunium, americium, and curium.

Cathode processor A high-temperature vacuum distillation furnace used to separate salt from metallic actinides deposited on an electrolyzer cathode.

Centrifugal contactors Liquid–liquid extraction equipment used for aqueous solvent extraction that consists of a spinning rotor to intensely mix the different phases.

Ceramic waste The glass-bonded sodalite matrix used to encapsulate waste salt from electrolyzer operation.

COEXTM French process for coextracting uranium and plutonium using extraction methods similar to PUREX.

Electrorefiner An electrochemical system used to separate actinides from spent fuel using a molten salt electrolyte.

Experimental Breeder Reactor-II A sodium-cooled, fast test reactor operational at Argonne National Laboratory-West from 1963 to 1994.

Geologic repository A permanent nuclear waste disposal site located deep within a geological formation.

Metal waste The stainless steel–zirconium matrix used to encapsulate cladding hulls and noble metals left in anode baskets after U electrolyzing is completed.

Mixer-settler Liquid–liquid extraction equipment used for aqueous solvent extraction requiring a relatively large footprint.

Minor actinides Actinide elements other than U and Pu. In spent fuel, the primary minor actinides of concern are Np, Am, and Cm.

Noble metals Elements found in spent nuclear fuel that have chloride forms that are thermodynamically less stable than uranium chloride.

Pulsed columns Liquid–liquid extraction equipment used for aqueous solvent extraction involving a single column consisting of trays of perforated plates to promote interphase mass transport.

Pyroprocessing Nuclear fuel treatment technology that uses electrochemical reactors with molten salt electrolytes to separate actinides from fission products.

PUREX Nuclear reprocessing technology that separates actinides from the spent fuel via liquid–liquid extraction involving acidic aqueous and organic liquid phases.

Spent fuel Nuclear fuel that has gone through at least one irradiation cycle in a nuclear reactor. It contains a mixture of actinides and fission products.

Solvent extraction A separations method for extracting species from a liquid phase. In this entry, it refers to a process for removing uranium from spent fuel that utilizes dissolution in acid followed by liquid–liquid extraction between aqueous and organic liquid phases.

UREX A variant of the PUREX process that separates uranium from spent fuel without recovering pure plutonium

V-blender A v-shaped vessel that is designed to efficiently blend two or more different kinds of powders with or without applied heat.

Objective

The objective of this entry is to give a basic overview of the technology elements behind nuclear fuel reprocessing. It should serve as a starting point for more detailed study with the aid of the Bibliography section to obtain more technical details on this subject. Several more process concepts have been proposed, tested, and demonstrated other than those listed in this entry. For the sake of conciseness, only two

fundamentally different technologies have been described here – aqueous and pyrochemical fuel reprocessing. In the case of pyrochemical fuel reprocessing, focus has been placed on the LiCl-KCl electrorefining technology developed originally at Argonne National Laboratory. The overall scope of nuclear fuel reprocessing technology is too broad to cover in this entry.

Introduction

Nuclear reprocessing is the chemical treatment of spent fuel involving separation of its various constituents. Principally, it is used to recover useful actinides from the spent fuel. Radioactive waste that cannot be reused is separated into streams for consolidation into waste forms. The first known application of nuclear reprocessing was within the Manhattan Project to recover material for nuclear weapons. Currently, reprocessing has a peaceful application in the nuclear fuel cycle. A variety of chemical methods have been proposed and demonstrated for reprocessing of nuclear fuel. The two most widely investigated and implemented methods are generally referred to as aqueous reprocessing and pyroprocessing. Each of these technologies is described in detail in [Sect. 3](#) with numerous references to published articles.

Reprocessing of nuclear fuel as part of a fuel cycle can be used both to recover fissionable actinides (primarily U and Pu isotopes) and to stabilize radioactive fission products into durable waste forms. It can also be used as part of a breeder reactor fuel cycle that could result in an almost 70-fold increase in energy utilization per unit of natural uranium [1]. Reprocessing can also impact the need for geologic repositories for spent fuel. The volume of waste that needs to be sent to such a repository can be reduced by first subjecting the spent fuel to reprocessing. The extent to which volume reduction can occur is currently under study by the US Department of Energy via research at various national laboratories and universities. Reprocessing can also separate fissile and nonfissile radioactive elements for transmutation.

The current known reserves of uranium that can be economically harvested are 5.5 million metric tons U at a maximum market price of about \$80/lb. At the current usage rate of 65,000 metric tons U/year, this

quantity of uranium will last for about 85 years [2]. However, nuclear power expansion in India, China, and other countries will soon lead to a substantial increase in the global usage rate for uranium. Increased investment in uranium exploration will undoubtedly reveal additional recoverable resources. And the inevitable increase in cost of uranium will lead to a higher fraction of economically recoverable resources. Reprocessing of spent fuel and use in light water reactors can also serve to improve efficiency of uranium resource utilization. Estimates range from about 10% to 30% for reduction of natural uranium usage as a result of reprocessing spent fuel and reuse in nonbreeding reactors.

Reprocessing Technology

Aqueous Reprocessing

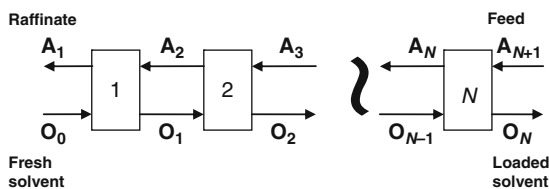
General Description With this technology, nuclear fuel is dissolved into an acidic solution. The resulting solution is then chemically processed to separate the metals of interest, typically uranium and/or plutonium. Minor actinides as well as other fission products may also be separated using advanced aqueous processes. Specific unit operations utilized in the past primarily consist of precipitation and liquid–liquid extraction. Precipitation was the primary method used in the defense industry initially for the production of plutonium. Liquid–liquid extraction was later developed as an improved method for reprocessing in the defense industry and also became the primary method of reprocessing of commercial nuclear fuel internationally. Since the vast majority of aqueous reprocessing applications continue to utilize liquid–liquid extraction, this will be the primary focus of discussion in this entry.

Liquid–liquid extraction (also called solvent extraction) was initially utilized in the petroleum industry beginning in the 1930s. It has since been utilized in numerous applications including petroleum, hydrometallurgical, pharmaceutical, and nuclear industries. Liquid–liquid extraction describes a method for separating components of a solution by exploiting an unequal distribution of the component(s) between two immiscible liquid phases. In most cases, this process is carried out by intimately mixing the two immiscible phases, allowing for the selective transfer

of solute(s) from one phase to the other, then allowing the two phases to separate. Typically, one phase will be an aqueous solution, usually containing the component(s) to be separated, and the other phase will be an organic solvent, which has a high affinity for some specific components of the solution. The process is reversible by contacting the solvent loaded with solute(s) with another immiscible phase that has a higher affinity for the solute than the organic phase. The transfer of solute from one phase into the solvent phase is referred to as extraction and the transfer of the solute from the solvent back to the second (aqueous) phase is referred to as back-extraction or stripping. The two immiscible fluids must be capable of rapidly separating after being mixed together, and this is primarily a function of the difference in densities between the two phases.

While limited mass transfer can be completed in a single, batch equilibrium contact of the two phases, one of the primary advantages of liquid–liquid extraction processes is the ability to operate in a continuous, multi-stage countercurrent mode. This allows for very high separation factors while operating at high processing rates. Countercurrent operation is achieved by repeating single-stage contacts, with the aqueous and organic streams moving in opposite directions as shown in Fig. 1.

In this flow diagram, the aqueous feed stream containing the solute(s) to be extracted enters at one end of the process (A_{N+1}), and the fresh solvent (organic) stream enters at the other end (O_0). The aqueous and organic streams flow countercurrently from stage to stage, and the final products are the solvent loaded with the solute(s), O_N , leaving stage N and the aqueous raffinate, A_1 , depleted in solute(s) and leaving stage 1. In this manner, the concentration gradient in the process remains relatively constant.



Nuclear Fuel, Reprocessing of. Figure 1
Countercurrent multistage extraction process flow diagram

The organic at stage O_0 contains no solute(s), while the raffinate stream is depleted of solute(s). Streams A_{N+1} and O_N contain the highest concentration of the solute(s).

For the process to be economical, the solvent must be recycled. In order to recycle the solvent, the solute is subsequently stripped from the solvent, and the solvent is then recycled back to the countercurrent extraction process. This allows the solvent to be recycled indefinitely, until it has degraded (due to acid hydrolysis or radiolytic degradation) or the solvent composition has changed due to solubility in the aqueous phase and/or evaporation.

History of Aqueous Separation Technology Aqueous separations processes for nuclear reprocessing evolved from early US defense programs for the separation of Pu for weapons manufacture. The bismuth phosphate process began operation at the Hanford Nuclear Reservation in 1944 for the separation of Pu from irradiated slugs from the B reactor [3]. The first continuous solvent extraction reprocessing plant replaced the bismuth phosphate process at Hanford in 1952. This facility used the reduction oxidation (REDOX) process to separate Pu. The REDOX process utilized methyl isobutyl ketone as an extractant. In the REDOX process, uranium and plutonium nitrate is preferentially extracted from the fission products in a high salting strength aqueous solution [4, 5]. The uranium and plutonium are then selectively stripped from the solvent by adjusting the valence state of the Pu to back-extract it and use a low salting strength strip solution to back-extract U. Additional cycles of extraction were used to decontaminate the products. General Electric's Knolls Atomic Power Laboratory developed the PUREX process in the 1950s. PUREX-based operations at Savannah River F Canyon began in 1954 and replaced the REDOX process at Hanford in 1956 [3]. The PUREX process became the standard method of reprocessing used nuclear fuel throughout the world.

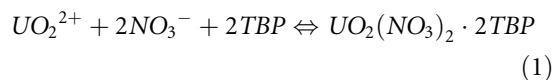
PUREX Process Technology PUREX-based reprocessing consists of leaching the spent fuel from the cladding using a nitric acid solution, chemical adjustment and filtration of the resulting feed solution, several cycles of solvent extraction to separate and purify the uranium and plutonium, solidification of the resulting uranium

and/or plutonium product, as well as the waste solutions. The plutonium oxide product, with or without uranium, is then recycled as mixed oxide (MOX) fuel. This resulting MOX fuel can be used as an alternative to low-enriched uranium in light water reactors. MOX fuel is widely used in Europe, and there are plans for use in Japan. About 40 reactors in Europe (Belgium, Switzerland, Germany, and France) are licensed to use MOX fuel [6]. Existing aqueous commercial reprocessing facilities throughout the world utilize, primarily, the PUREX solvent extraction process or a variant of this process to accomplish the separation of U and Pu. Pure plutonium can also be separated with this process and used to make nuclear weapons.

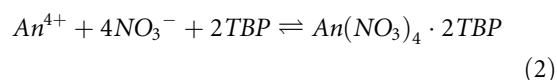
The front end of the PUREX process involves mechanical chopping of the spent nuclear fuel assemblies into small pieces (1 cm long) followed by leaching of the spent fuel in a nitric acid solution. The chopped pieces of the pins, as well as spacers and other fittings, must then be separated from the leached fuel solution. This has been performed through the use of perforated baskets that hold the hardware, such as in the batch operations performed at the THORP facility in the UK, or through the use of a continuous dissolver, such as the wheel dissolver at the UP2 and UP3 plants in France which holds the hardware in buckets formed in sections of the wheel as the wheel rotates through the nitric acid solution [7]. The dissolver solution, after further clarification and feed adjustment, is then processed through the use of the PUREX technology to separate and purify the uranium and plutonium from the dissolver product solution.

The PUREX process utilizes 20–40 vol% tributyl phosphate in a hydrocarbon diluent to extract uranium

and plutonium from the acidic solution resulting from the dissolution of spent nuclear fuel. In general, metals in the +4 and +6 oxidation state are extracted in the PUREX process. The chemical equilibria for U (VI) in a nitrate media is

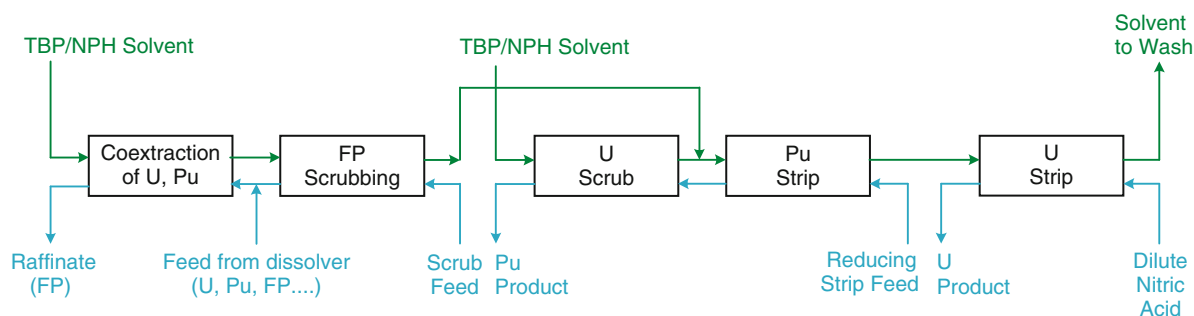


The chemical equilibria for the actinides in the +4 oxidation state is



Plutonium maintained in the +4 oxidation state is, therefore, coextracted with the U. Pu (III) and other actinides and lanthanides in the +3 or lower oxidation state are not extracted in the PUREX Process. Neptunium, if maintained in the +6 oxidation state, can be coextracted with the U and Pu. The strong extraction of the +4 and +6 oxidations states, along with the weak extraction of the other oxidations states, results in the effective use of the PUREX process for separation of uranium and plutonium from nearly all of the other metals present in the spent nuclear fuel.

A typical PUREX process first extraction cycle is provided in Fig. 2. The solution resulting from the dissolution of the spent nuclear fuel is the feed to the coextraction section of the flowsheet. The aqueous feed flows countercurrent to the PUREX solvent, and the U and Pu are extracted by the TBP into the normal paraffin hydrocarbon (NPH) organic phase. The loaded organic phase enters the fission product (FP) scrub section in which a nitric acid scrub



Nuclear Fuel, Reprocessing of. Figure 2
Typical PUREX first cycle flowsheet

solution (approx. 2 M HNO_3) is used to remove coextracted fission products, such as Zr and Ru, from the solvent. The scrub solution containing the Zr and Ru combines with the feed solution entering the extraction section. The solvent then enters a Pu strip section in which the Pu is back-extracted from the organic phase. This is accomplished by reducing the Pu from the extractable +4 oxidation state to the inextractable +3 state. A strip solution containing a reductant, such as hydroxylamine nitrate, U (IV), or ferrous sulfamate, is typically used [8].

Reduction and back-extraction of the Pu also results in back-extraction of a portion of the uranium. The strip product from the Pu strip section therefore enters a uranium scrub section in which the Pu strip solution is contacted with a fresh solvent feed to reextract this uranium into the organic phase. The organic phase containing the reextracted U combines with the loaded solvent from the extraction section which enters the Pu strip section. Once the Pu has been back-extracted from the PUREX solvent, the solvent enters the uranium strip section, which utilizes dilute nitric acid (typically 0.01 M HNO_3) at elevated temperature to back-extract the U into the aqueous phase.

The resulting solutions from the first cycle PUREX extraction process include a solvent solution that is washed with a carbonate or hydroxide solution to remove degradation products and recycled back to the extraction section, a raffinate stream which is depleted of the U and Pu and disposed of as waste, and the Pu and U product streams. The U and Pu product streams are typically further processed with additional PUREX cycles to purify these streams [9].

PUREX Process Waste Treatment The separation of the Pu and U from spent nuclear fuel results in a high-level waste (HLW) requiring immobilization and storage. The immobilization technology currently in use in the UK, France, and Japan is vitrification of the waste to form a stable borosilicate glass waste form suitable for long-term storage [10]. The glass waste produced is poured into canisters and are stored until long-term geological storage is available. Appropriate geological repositories are currently being pursued in these countries. Low (LLW) and intermediate-level wastes (ILW) are also generated from aqueous reprocessing and require treatment and disposal.

Advanced Aqueous Separations Process Technology

Advanced aqueous separations processes are being developed throughout the world including the USA, France, UK, Japan, Russia, and China. The goals of the development of advanced aqueous processes include closing the nuclear fuel cycle and reducing the proliferation risk of the technologies. Reduction in proliferation risk is being addressed through development of modified PUREX processes which do not separate pure plutonium. In addition, advanced separation processes are being developed to separate the actinides for reactor recycle to close the fuel cycle.

Advanced Aqueous Reprocessing Strategies

Numerous strategies are being developed internationally for advanced aqueous reprocessing processes. The goals of these processes typically are to accomplish separations beyond the Pu and U that is separated with the PUREX process in order to reduce the volume, radiotoxicity, and heat generation of the spent nuclear fuel. The components targeted include the minor actinides as a group, the short-lived fission products (Cs and Sr), and/or individual actinides, such as Am. The minor actinides or Am separated from the spent fuel would be recycled for burning in a fast reactor. If separated, the Cs and Sr could be placed into decay storage.

A wide variety of advanced processes are currently being developed to accomplish these advanced separations. Major research efforts on advanced processing are ongoing in France, Japan, UK, USA, China, and Russia [11–17]. These technologies are at various stages of maturity, and none have been implemented into full-scale reprocessing facilities to date.

Advanced aqueous processing has the potential to significantly reduce the volume, heat load (long and short-term), and radiotoxicity of HLW requiring disposal in a geological repository [18]. These processes, however, will result in the generation of a significant quantity of LLW requiring treatment and disposal. These wastes include spent solvent, solvent treatment solutions, and decontamination solutions, among others. Additionally, solid waste is generated from facility operations (e.g., gloves, shoe covers, cleaning supplies, filters) and will require treatment and disposal.

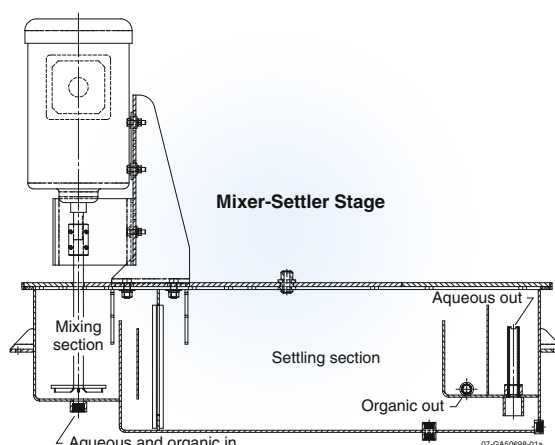
The primary focus of recent development of advanced PUREX processes is to prevent the separation of pure Pu, thus reducing proliferation risk, as well as

controlling Np and Tc chemistry to allow for the extraction of these metals.

In France, the COEX™ process is being developed which coextracts the U and Pu and produces a Pu/U product instead of a pure Pu product [19]. This is accomplished by adjusting the chemistry of the PUREX process to allow some U to be back-extracted from the solvent with the Pu. The process also uses a coconversion process to produce a (Pu, U) oxide product. Coextraction of Np is also being evaluated with the COEX™ process to produce a (Pu, U, Np) oxide product [19]. The Rokkasho reprocessing plant in Japan, which has been constructed and is currently undergoing operational testing, also uses a PUREX process that has been modified to combine uranium with the separated plutonium in a 50/50 mix prior to denitration [20]. In the USA, the uranium extraction (UREX) process has been developed which separates the uranium from spent nuclear fuel. The UREX process is a modification of the PUREX process in which the Pu is prevented from extracting with the U by adding acetohydroxamic acid as a reductant/complexant [21, 22].

Aqueous Reprocessing Equipment The solvent extraction equipment utilized for industrial-scale aqueous reprocessing must enable continuous processing at high throughputs while efficiently mixing and separating the two phases. In the nuclear industry, specific constraints, such as remote operation and maintenance, must be considered, since the solutions processed are highly radioactive. There are three basic types of equipment used in industrial-scale nuclear solvent extraction processes: mixer-settlers, columns, and centrifugal contactors. A detailed description of the three types of equipment follows.

Mixer-Settlers. This equipment consists of a small mixing chamber followed by a larger gravity-settling chamber as shown in Fig. 3. Each mixer-settler unit provides a single stage of extraction. The two phases enter the mixing section where they are mixed using an impeller. The two-phase solution flows into the settling section where the phases are allowed to separate by gravity due to their density differences. Typical mixer settlers have mixing times on the order of a few minutes and settling times of several minutes. The separate phases exit the settling section by flowing over a weir



Nuclear Fuel, Reprocessing of. Figure 3
Diagram of a mixer-settler

(less dense phase – typically organic) or through an underflow then over a weir (more dense phase – typically aqueous). The separation interface is controlled by the height of the weirs on the outlets of the settler section. Only minimal instrumentation is required, and mechanical maintenance is limited to occasional mixing motor replacement. In a counter-current process, multiple mixer settlers are installed with mixing and settling chambers located at alternating ends for each stage (since the outlet of the settling sections feed the inlets of the adjacent stage's mixing sections).

Mixer-settlers are best suited for use when a process requires longer residence times and when the solutions are easily separated by gravity. They require a large facility footprint, but do not require much headspace and need limited remote maintenance capability for occasional replacement of mixing motors [23, 24].

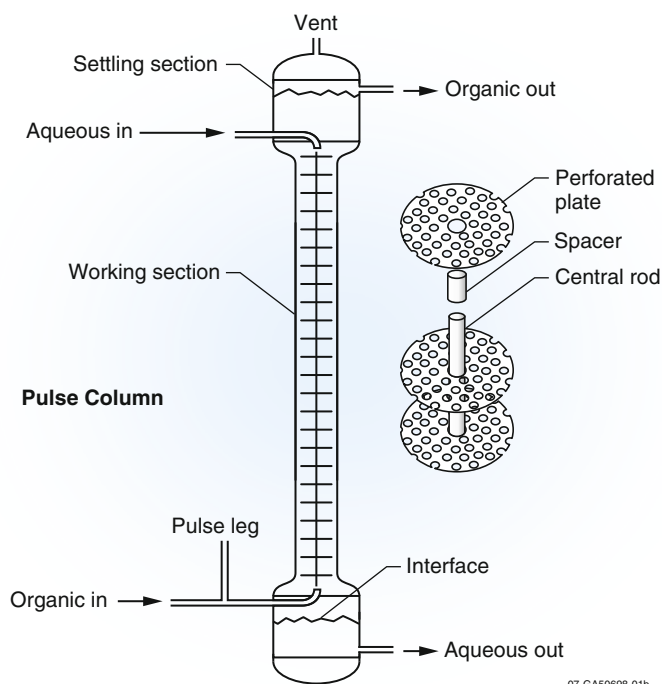
Columns. There are two basic types of columns employed industrially, packed columns and pulse columns with plates or trays. Packed columns are filled with some type of packing material, such as Raschig rings, to create a tortuous path for the two solutions as they flow through the column (typically aqueous feed downward and solvent upward), ensuring that the two phases are in constant contact. Packed columns have no moving parts and are relatively simple to operate, but they are not very efficient. Since columns do not have

discrete stages, such as mixer-settlers or centrifugal contactors, the number of stages is determined by the height equivalent of a theoretical stage (HETS) [25]. For most packed columns, this HETS of extraction is usually several feet, meaning that a countercurrent process utilizing several stages to effect a given separation factor would require very tall columns. To reduce the height of a theoretical stage in the column, other packing (trays or perforated plates) is used and mechanical energy is applied to force the dispersed phase into smaller droplets, improving mass transfer. The most common type of column used today, particularly in the nuclear industry, is the pulse column.

In a pulse column, liquids are continuously fed to the column and flow counter – currently, as is done with a packed column; but mechanical energy is applied to pulse the liquids in the column up and down. This is normally done by injecting pressurized air into a pulse leg that pushes liquid into the column, then venting the pulse leg to fill the pulse leg with

solution from the column. The pulse action lifts and lowers the solution in the column, usually only a few inches. The column is filled with perforated plates or other plates to promote droplet formation as the dispersed phase is pushed through the plates. This pulsing action reduces droplet size of the dispersed phase and improves mass transfer. A perforated plate pulse column is shown in Fig. 4. There are a number of plate designs used. Early pulse columns used sieve plates, which are flat plates with holes drilled into them. A more effective plate is the nozzle plate, which has different contours on the top and bottom of the plate (making it directional, in that, it must be configured according to the continuous phase in the column). The French and Japanese pulse columns employ a disk and doughnut configuration, where the plates are solid (no openings) but the alternating plates enable effective contacting of the phases [25].

The separation interface is controlled during column operation using bubble probes in the disengaging



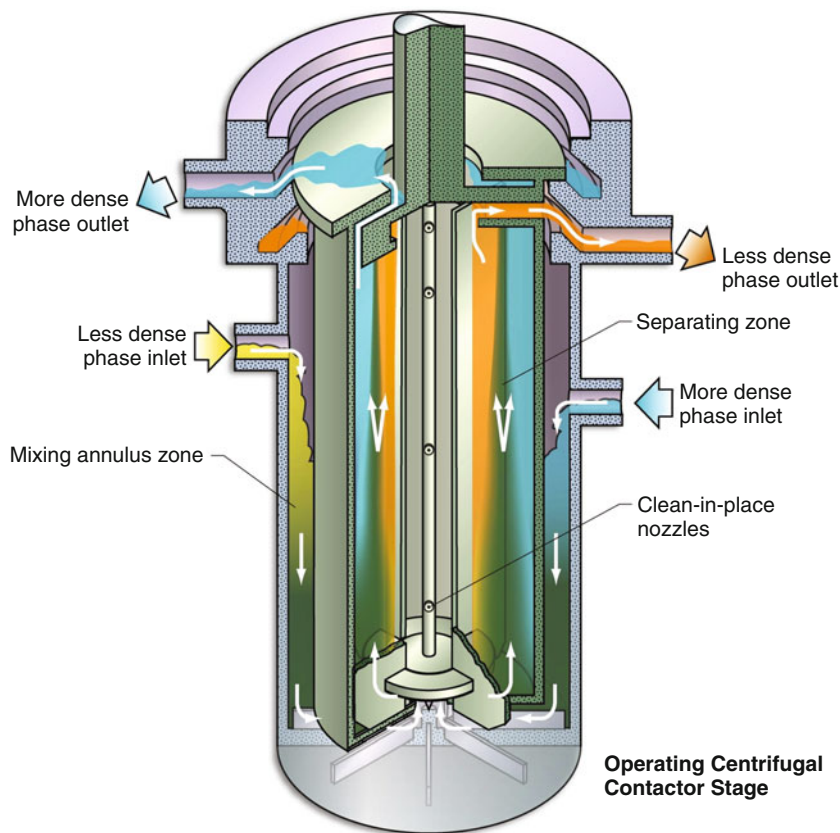
07-GA50698-01b

Nuclear Fuel, Reprocessing of. Figure 4
Pulse column with perforated plates

section. The probes are interfaced to a controller that drains heavy phase from the bottom of the column. The bubble probes allow operators to monitor the weight of the column, which gives them a good indication of column performance, by determining the ratio of heavy and light phases in the column. In addition, pulsing devices and pulse speed controllers are required as pulse frequency and amplitude must be controlled during operation. Periodic maintenance is required only for the pulsing equipment, which is located out of cell, above the columns. Pulse columns are used when a process requires intermediate residence times, as adjusting flow rate easily varies residence time. They require a small facility footprint, but do require much headspace (typically 40–50 ft). Pulse columns do not need remote maintenance capability, as all moving parts (pulsar equipment) are located outside the shielded cell. Extensive literature has been

published on pulse columns [9, 27–29]. Pulse columns are the primary type of aqueous separation equipment utilized in the nuclear industry today.

Centrifugal Contactors. Centrifugal contactors, like mixer-settlers, are discrete-stage units, providing one stage of extraction per unit and are readily linked together as each rotor pumps separated fluids to the next stage inlet in each direction. The primary difference between a centrifugal contactor and a mixer-settler is the separation of the two-phase mixture. Centrifugal contactors employ a spinning rotor that intensely mixes the two phases and separates the two phases inside the rotor where the centrifugal forces can be as high as 300 g. This results in efficient and fast phase separation. The separated phases exit the contactor by overflow and underflow weirs, similar to a mixer-settler. A cutaway view of an operating centrifugal contactor is shown in Fig. 5.



07-GA50698-01c

Nuclear Fuel, Reprocessing of. Figure 5
Cutaway view of an operating centrifugal contactor

Centrifugal contactors have high single-stage efficiency (routinely greater than 95% of theoretical for chemical processes with rapid kinetics). Process flow interruptions cause no loss of process concentration profiles if centrifugal contactor rotors are kept spinning. Thus centrifugal contactor-based processes can be paused for a period of time sufficient to reestablish flow or even replace a motor without significant loss of product or rework. Centrifugal contactors require minimal instrumentation for process operation. Computer control via commercial software allows monitoring of motor amperage, rotor rpm, inlet flow rates, temperatures, and many other process parameters. Centrifugal contactors are used when a process requires short residence times, on the order of several seconds. They require a small facility footprint and minimal headspace, but they do require remote maintenance capability for periodic removal of the motor and/or rotor.

Centrifugal contactors have been the subject of much recent development work over the past 40 years, while the designs of pulse column and mixer-settlers has changed little over the same time period [30–32]. Early designs included a paddle wheel to mix the phases below the spinning rotor [33]. This precluded removal of the rotor assembly. The annular centrifugal contactor was subsequently developed, which allowed the motor and rotor assembly to be easily removed [34]. Other designs included multistage units, units for low-mix applications (higher phase separation), and clean-in-place units that have an array of internal spray nozzles to facilitate solids removal as necessary [35–37]. Design of remote operation and maintenance capabilities has also continued, resulting in more efficient remote handling [38, 39].

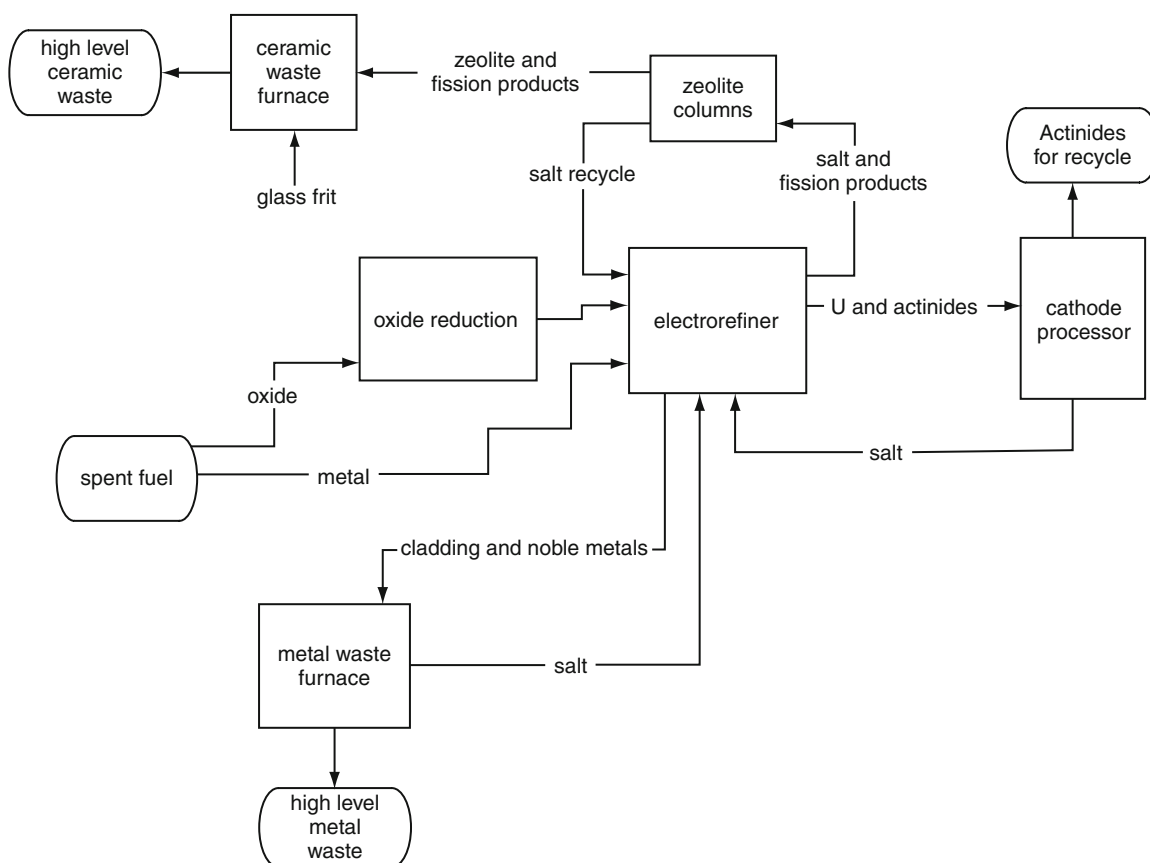
Pyroprocessing

General Description Pyroprocessing is currently considered an alternative reprocessing technology to the more commonly used aqueous processing technology. Pyroprocessing accomplishes separations by way of high-temperature electrorefining. It is yet to be implemented on a large scale, limited to date to laboratory-scale and engineering-scale experimentation and demonstration. Much of the current state of the art for pyroprocessing was developed during the Integral Fast Reactor (IFR) program, which was carried out

at Argonne National Laboratory from about 1984 to 1995 [40, 41]. With the shutdown of Experimental Breeder Reactor-II in 1995, the IFR program was converted into a spent fuel treatment program to safely treat the 25 metric tons of heavy metal from that reactor [42]. Pyroprocessing utilizes molten salt electrolytes as the media rather than acidic aqueous solutions and organic solvents [43]. These electrolytes are principally used to support electrochemical separations such as uranium electrorefining and electrolytic reduction of oxide fuel. The process includes vacuum furnaces that accomplish salt/metal separations and melt metal deposits into ingots for either waste disposal or fuel fabrication. Ceramic and metal waste streams are generated that immobilize fission products and, optionally, plutonium and minor actinides into high-level waste forms. For eventual commercial implementation, it is expected that plutonium and minor actinides will be recycled and used for fast reactor fuel fabrication. While this technology is yet to reach the commercialization stage, it has been the subject of extensive, government-funded research and development worldwide in addition to the EBR-II spent fuel treatment work in the USA. For example, the Republic of Korea is currently pursuing a strategy of developing pyroprocessing technology for treatment of spent fuel from their commercial light water reactors to minimize volume of HLW and possibly extract fissile actinides for eventual fabrication of fast reactor fuel [44, 45]. Russia has already demonstrated production of MOX based on pyroprocessing and plans to develop a closed fuel cycle using the technology by 2020.

While PUREX and related aqueous reprocessing technology has superior maturity, pyroprocessing does have unique benefits that make it a credible alternative and in some cases a preferred alternative. This includes use of process liquids that are more stable than organics in the presence of high radiation fields, improved criticality safety due to the lack of neutron moderators in the process, and waste processing that is integrated with the separations flowsheet.

Process Technology There are many variants of the pyroprocessing flowsheet, but the IFR scheme shown in Fig. 6 can be used as a reference, as it contains all of the key unit operations.



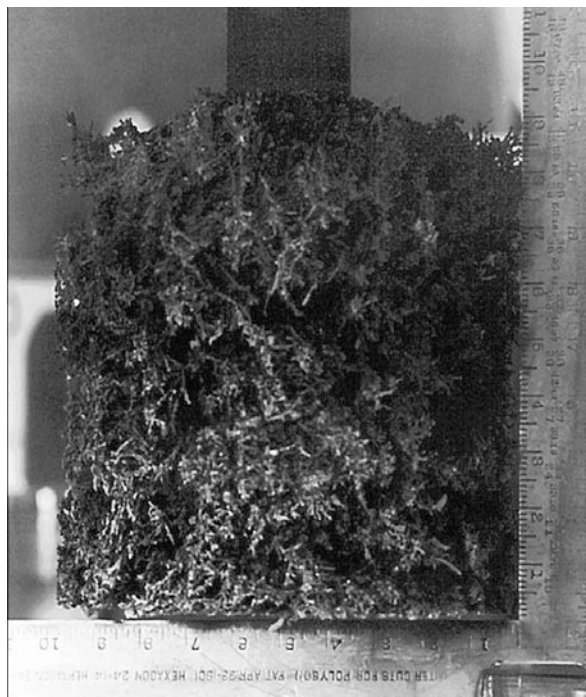
Nuclear Fuel, Reprocessing of. Figure 6

Fuel Processing Flowsheet for the Integral Fast Reactor Program

The electrorefiner is at the center of the flowsheet and is used to perform the primary separation of actinides from fission products [46, 47]. It contains a molten salt electrolyte – typically LiCl-KCl-UCl_3 maintained at 450–500°C. The eutectic composition of LiCl-KCl (44.2 wt% LiCl , 55.8 wt% KCl) is maintained to keep the melting point at approximately 350°C. The UCl_3 content varies depending on desired operating conditions from about 0.5–10 wt%. It is used as a charge carrier for electrotransport through the electrolyte. After the spent fuel is chopped into segments, it is loaded into anode baskets, and the baskets are lowered into the electrorefiner. As current is passed between the anode and cathode, U metal is oxidized to U^{3+} at the anode and reduced back to metallic form at the steel cathode. The deposit contains high purity uranium and is typically dendritic. An example of a U cathode deposit is shown in Fig. 7.

Transuranic (TRU) elements and active metal fission products are oxidized electrochemically or via reaction with uranium chloride in the salt and enter the electrolyte. Under normal conditions, Pu and minor actinides cannot deposit at the cathode, because their back-reaction with UCl_3 is thermodynamically spontaneous. However, codeposition of U and TRU can be achieved via a combination of elevating the TRU to U ratio in the salt and utilizing a liquid cadmium cathode (LCC). In the molten cadmium phase, TRU elements have a very low activity coefficient compared to U. This allows for TRU metals to be present in quantities comparable to that of uranium in the cadmium. Alternative methods are currently being investigated to corecover U and TRU without the need for an LCC.

Fission product elements segregate between the anode basket and the molten salt during the



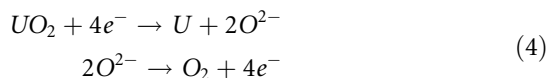
Nuclear Fuel, Reprocessing of. Figure 7
Dendritic uranium deposit on electrorefiner cathode

electrorefining process. Noble metals such as Tc, Ru, and Rh remain with the cladding hulls in the anode basket. Active metals that typically include Group I and II elements in addition to lanthanides are oxidized to chloride form and accumulate in the salt. If sodium metal is used as a bonding agent, as in the case of EBR-II fuel, this sodium is oxidized to sodium chloride, which accumulates in the ER electrolyte.

Note in Fig. 6 that both metal fuel and oxide fuel can be treated via pyroprocessing. Treatment of metallic fuel is relatively straightforward due to the fact that it is already in a state compatible with the ER. Oxide fuel must first be converted to metallic form. This can be accomplished in an oxide reduction step. Various methods have been investigated for reducing spent oxide fuel. Early efforts were focused on chemical reduction via lithium. Similar to electrorefining, a molten salt is used for carrying out this reaction. In this case, it is LiCl saturated with lithium metal at 650°C. The lithium reduction reaction is as follows [47].



More recently, a similar process based on electrochemical reactions has been favored for development. It also uses a vessel containing molten LiCl at 650°C, but it contains lithium oxide in the salt rather than lithium metal. The reactions for the electrolytic process are shown below [49].



The generated oxygen bubbles out of the salt as a gas and can be sent to an off-gas treatment system to remove any entrained or volatile contaminants. The advantages of the electrolytic method based on the above reactions are that lithium oxide concentration in the molten salt can be kept low (1 wt%), and there is no need for a separate vessel to regenerate lithium metal from lithium oxide.

U or U/TRU product deposited on the cathode in the electrorefiner is transferred to a cathode processor, which is essentially a vacuum distillation furnace. The salt is separated from the metals and recycled to the electrorefiner. The purified metals can be fed into a process for fabricating metallic fuel for fast reactors. For the case of the EBR-II Spent Fuel Treatment process, the cathode processor operates at a temperature up to 1200°C and achieves pressures less than 1 torr.

After an electrorefining run, the anode basket contains the cladding hulls, undissolved actinides, inert fuel matrix material such as zirconium, adhering salt, and noble metal fission products such as Tc, Mo, Rh, and Ru. All of this material is removed from the anode baskets and loaded into an inductively heated vacuum distillation furnace that is used to distill adhering salt and consolidate the metals into an ingot. The metal ingot becomes a waste form that has been tested and shown to be suitable for disposal in an HLW repository [50, 51].

Electrorefiner salt becomes progressively more contaminated with fission product chlorides as well as sodium chloride in the process of treating fuel. Once the contamination level has exceeded a predetermined limit, the salt must be removed from the electrorefiner and either disposed or processed through a purification step and returned to the electrorefiner. The basis for that limit can be fission product decay heat, salt melting point, or criticality limits. Another potential limiting factor is contamination of the metallic actinide

products recovered in the cathode processor. High concentrations of rare earth fission products in the salt, for example, have been shown to lead to high rare earth contamination levels in the actinide product. The process flow sheet shown in Fig. 6 includes zeolite ion exchange columns for achieving this salt purification. Zeolite-A has been shown to exhibit selectivity for the fission product ions when in contact with molten chloride salt [52–55]. Other alternatives that have been considered for treating the salt to remove fission products and other contaminants include selective precipitation, zone freeze refining, and adsorption by nonzeolitic materials [56].

The current baseline technology for dealing with the salt waste from electrorefining EBR-II spent fuel is to nonselectively immobilize the salt into a ceramic waste form consisting of glass-bonded sodalite [57, 58]. In this process, salt is removed from the electrorefiner, sized via crushing and milling to a fine particulate, and absorbed into zeolite-4A in a high-temperature blending operation. A V-blender capable of heating and mixing particulate material to 500°C is used for this absorption step. Prior to being loaded into the V-blender, it is necessary to dry the zeolite to less than 1 wt% water. This drying is used to maximize salt absorption in the zeolite while minimizing evolution of water vapor in a high-temperature, corrosive environment. Drying the zeolite should also minimize pores in the final ceramic waste form. Zeolite drying is accomplished via mechanically fluidizing the zeolite under vacuum at temperatures up to 550°C [57, 58]. Heating the zeolite-4A to temperatures of 600°C or higher has been determined to cause structural damage that inhibits its ability to absorb salt [59]. Final consolidation into the ceramic waste form occurs after borosilicate glass binder has been mixed with the salt-loaded zeolite, loaded into a steel canister, and heated to a maximum temperature of 915–950°C. During the process of consolidation, the zeolite-A phase converts to sodalite. In the glass-bonded sodalite waste form, the fission products are distributed between the glass and sodalite phases [60].

If an ion exchange process with zeolite-A has been used to selectively remove fission products from the salt, the resulting fission product-loaded zeolite-A can be similarly converted into a glass-bonded sodalite ceramic waste form. Zeolite-A used for ion exchange is typically

in pelletized form and must be milled to a fine particulate prior to blending with additional dried zeolite-4A and borosilicate glass. The flowsheet shown in Fig. 6 includes zeolite ion exchange followed by conversion of this zeolite into the ceramic waste form.

Future Directions

In the USA, the current focus is on research and development into both aqueous reprocessing and pyroprocessing technology to support a future decision on closing the fuel cycle. The US Department of Energy (2009) has established the Fuel Cycle Research and Development program for carrying out this research in national laboratories and universities. At this time, there are no large-scale demonstration projects planned. Meanwhile, plans to open a geologic repository for spent nuclear fuel and waste in Nevada's Yucca Mountain have been suspended. The government has commissioned a study to evaluate alternative options for disposal of the spent fuel and waste.

In Japan, the main option for reprocessing spent nuclear fuel is based on aqueous process technology. The Rokkasho plant based on such technology is currently operational with a design capacity of 800 tons of spent light water reactor fuel per year, extracting up to 8 tons of plutonium per year for MOX fuel production. Pyroprocessing is considered an option for fast reactors once they have been included in the Japanese energy fleet. A commercial fast reactor is not planned for completion in Japan until about 2050.

In France, advanced aqueous processing technologies are being developed and assessed to support future recycling of Am and Cm or the minor actinides together with U and Pu to fast reactors. Pilot-scale demonstration is planned within the next decade with a goal of industrial deployment to support the deployment of Generation IV fast reactors. France is also in the process of selecting a site for a geological repository for disposal of HLW with a goal to open the repository in 2025. Research related to the study of geological formations and the capacity as a deep geological repository for HLW is being conducted at the Meuse/Haute Marne Underground Research Laboratory located in Bure, France.

Russia is currently reprocessing spent fuel from civilian power reactors as well as spent HEU fuel from naval and other reactors at Mayak's RT-1 aqueous

reprocessing plant. The Experimental-Demonstration Center (EDC), which will be a 100 metric tons/year pilot facility for evaluation of the fuel cycle based on modified PUREX extraction technology, is currently being designed. This facility will also be used to develop other advanced processing technologies for processing used fuel from thermal reactors. The current goal is to support completion of a new aqueous reprocessing facility around 2025. Research is also actively being performed relative to pyroprocessing technologies for the processing of spent fuel from future fast reactors. To this end, the Multipurpose Pyroprocessing Complex (MPC) is being designed at the Research Institute of Atomic Reactors (RIAR) to support molten salt processing development at a capacity of up to 2,500 kg fast reactor used fuel per year.

In South Korea (Republic of Korea), on-site wet storage capacity for spent nuclear fuel at its twenty operating nuclear power plants is rapidly approaching current limits. In December 2008, Atomic Energy Commission of South Korea decided to develop a closed fuel cycle associated with pyroprocessing and sodium fast reactors (SFR) with metallic fuels. A demonstration SFR is planned to operate from 2030 initially with U-Zr metal fuels and later with recycled U/TRU/Zr metal fuels produced from pressurized water reactor (PWR) spent nuclear fuel in a pyroprocessing facility that is planned to be operated from 2025. Pyroprocessing technology research and development continues at Korea Atomic Energy Research Institute with plans to build an engineering scale facility by 2016. Aqueous reprocessing technology is currently not being actively studied and is not considered a candidate for commercialization in the Republic of Korea. The lack of a high-level waste repository is another problem faced by the country due to severely limited land resources. Waste minimization is, thus, a major objective with pyroprocessing technology research and development in South Korea.

Bibliography

1. Cochran RG, Tsoulfanidis N (1993) *The nuclear fuel cycle: analysis and management*, 2nd ed. American Nuclear Society, Washington, p. 214
2. OECD, IAEA (2008) *Uranium 2007: resources, production, and demand*. Nuclear Energy Agency, Washington, June 2008
3. Gray LW (1999) From separations to reconstitution – A short history of plutonium in the US and Russia. Lawrence Livermore National Laboratory, UCRL-JC-133802
4. Evans TF, Tomlinson RE (1954) Hot semiworks REDOX studies, Hanford Atomic Products Operations, HW-31767
5. (1951) REDOX technical manual, Hanford Works, HW-18700
6. Hore-Lacy I (2009) Mixed oxide fuel (MOX). World Nuclear Association (Content Partner); Cutler J Cleveland (Topic Editor). In: Cutler J (ed) *Encyclopedia of earth. Environmental Information Coalition, National Council for Science and the Environment*, Cleveland/Washington, DC
7. Denniss IS, Jeapes AP (2001) Reprocessing irradiated fuel. In: Wilson PD (ed) *The nuclear fuel cycle: from ore to wastes*. Oxford University Press, Oxford, p 120
8. Poczynajlo A (1988) Studies on reductive back extraction of plutonium in PUREX process. *J Radioanal Nucl Chem* 125(2):445–465
9. Long JT (1967) *Engineering for nuclear fuel reprocessing*. Gordon Breach Sci Publ, New York
10. Petitjean V, Fillet C, Boen R, Veyer C, Flament T (2002) Development of vitrification process and glass formulation for nuclear waste conditioning, *Proceedings of Waste Management 2002*. Tucson, AZ USA
11. *Spent Fuel Reprocessing Options* (2008) International Atomic Energy Administrations, IAEA-TECDOC-1587, Vienna, Austria 2008
12. Boullis B (2008) Future nuclear fuel cycles: prospects and challenges. In: Bruce Moyer (ed) *Solvent Extraction: fundamentals to industrial applications*, *Proceedings of ISEC 2008 International Solvent Extraction Conference*, vol 1., pp 29–42
13. Nash K (2008) Key features of the TALSPEAK and similar trivalent actinide-lanthanide partitioning processes. In: Bruce Moyer (ed) *Solvent extraction: fundamentals to industrial applications*, *Proceedings of ISEC 2008 International Solvent Extraction Conference*, vol 1., pp 511–519
14. Laidler J (2008) An overview of spent-fuel processing in the Global Nuclear Energy Partnership. In: Bruce Moyer (ed) *Solvent extraction: fundamentals to industrial applications*, *Proceedings of ISEC 2008 International Solvent Extraction Conference*, vol 1., pp 695–701
15. Riddle C, Baker J, Law J, McGrath C, Meikrantz D, Mincher B, Peterman D, Todd T (2005) Development of a novel solvent for the simultaneous separation of strontium and cesium from acidic solutions. *Solvent Extr Ion Exch* 23(3):449–461
16. Christiansen B, Apostolidis C, Carlos R, Courson O, Glatz JP, Malmbeck R, Pagliosa G, Römer K, Serrano-Purroy D (2004) Advanced aqueous reprocessing in P&T strategies: process demonstrations on genuine fuels and targets. *Radiochim Acta* 92:475–480
17. Miguiditchian M, Chareyre L, Hérés X, Hill C, Baron P, Masson M (2007) GANEX: adaptation of the DIAMEX-SANEX process for the group actinide separation, *Proceedings of GLOBAL 2007 Advanced Nuclear Fuel Cycles and Systems*. Bosie, Idaho

18. Wigeland R, Bauer T, Fanning T, Morris E (2006) Separations and transmutation criteria to improve utilization of a geologic repository. *Nuclear Technol* 154(1):95–106
19. Drain F, Emin JL, Voinche R, Baron P (2008) COEX process: cross-breeding between innovation and industrial experience. *Proceedings from Waste Management 2008*, Tucson, AZ
20. Katsuta T, Suzuki T (2009) Japan's spent fuel and plutonium management challenge. *Energy Policy* doi: 10.1016/j.enpol.2009.05.075
21. Pereira C, Vandegrift G, Regalbuto M, Bakel A, Bowers D, Gelis A, Hebden A, Maggos L (2007) Lab-scale demonstration of the UREX+1a process using spent fuel. *Proceedings from Waste Management 2007*, Tucson, AZ
22. Nuñez L, Vandegrift G (2000) Evaluation of hydroxamic acid in uranium extraction process: literature review, Argonne National Laboratory, ANL00/35.
23. Colven, TJ Jr, (1956) Mixer-Settler development-operating characteristics of a large-scale mixer-settler. *Savannah River Laboratory*, DP-140
24. Davidson JK, Shafer AC, Haas WO (1957) Application of Mixer-Settlers to the PUREX Process. In: *The Symposium on the Reprocessing of Irradiated Fuels*, Book 1. United States Atomic Energy Commission, TID-7534
25. Benedict M, Pigford TH, Levi HW (1981) *Nuclear chemical engineering*. McGraw-Hill, New York, p 210
26. Milot JF, Duhamet J, Gourdon C, Casamatta G (1990) Simulation of a pneumatically pulsed liquid-liquid extraction column. *The Chemical Engineering Journal* 45:111–122
27. Sege G, Woodfield FW (1954) *Chem Eng Progress* 50(8)
28. Geier RG (1954) Application of the Pulse Column to the PUREX Process. USACC, Report TID-7534
29. Richardson GL, Platt AM (1961) *Progress in nuclear energy, Series IV, Technology engineering and safety*, vol 4. Pergamon Press, New York
30. Leonard RA (1988) Recent advances in centrifugal contactor design. *Separation Sci Technol* 23:12–13
31. Jubin RT et al (1988) Developments in centrifugal contactor technology. Oak Ridge National Laboratory, ORNL/TM-10768
32. Meikrantz DH, et al (2001) Annular Centrifugal Contactors for Multiple Stage Extraction Processes. *Chem Eng Comm* 188: 115–127
33. Watts C (1977) Solvent Extraction Equipment Evaluation Study – Part 2. Battelle Northwest Laboratory, BNWL-2186 Pt. 2
34. Bernstein GL et al (1973) A high-capacity annular centrifugal contactor. *Nuclear Technol* 20
35. Drain F et al (2003) Forty years of experience with liquid-liquid extraction equipment in the nuclear industry. *Proceedings from Waste Management Conference 2003*, Tucson, AZ
36. Meikrantz DH et al (1996) Rotor sleeve for a centrifugal separator. U.S. Patent # 5,571,070
37. Macaluso LL, Meikrantz DH (1999) Self-cleaning rotor for a centrifugal separator. U.S. Patent # 5,908,376
38. Garn, TG, Meikrantz DH, Law JD (2008) Remote evaluation of a three-stage 5 cm annular centrifugal contactor remote module at the INL. Idaho National Laboratory, INL/EXT-08-13670
39. Meikrantz DH, Garn TG, Law JD, Macaluso LL (2009) Evaluation of a new remote handling design for high throughput annular centrifugal contactors. Idaho National Laboratory INL/EXT-09-16824
40. Chang YI (1989) The integral fast reactor. *Nuclear Technol* 188(2):129–138
41. Till CE, Chang YI, Hannum WH (1997) The integral fast reactor – an overview. *Prog Nuclear Energy* 31(1–2):3
42. Benedict RW (1997) EBR-II spent fuel treatment demonstration project. *Trans Amer Nuclear Soc* 77:75–76
43. Ackerman JP (1991) Chemical basis for pyrochemical reprocessing of nuclear fuel. *Industrial Eng Chem Res* 30(1):141–145
44. Lee SY et al (2007) A preliminary study on the safeguardability of a Korean Advanced Pyroprocessing Facility (KAPF). *Proceedings of Global 2007*, Boise, Idaho
45. Lee HS, Hur JM, Ahn DH, Kim IT, Lee JH (2009) Development of Pyroprocessing Technology at KAERI. *Proceedings of Global 2009*, Paris, France
46. Willitt JL, Miller WE, Battles JE (1992) Electrorefining of uranium and plutonium – a literature review. *J Nuclear Mater* 195(3):229–249
47. Goff KM, Benedict RW (2005) Electrorefining Experience for Pyrochemical Reprocessing of Spent EBR-II Fuel. *Proceedings of Global 2005*, Tsukuba, Ibaraki (Japan)
48. Karell EJ, Gourishankar KV, Smith JL, Chow LS, Redey L (2001) Separation of actinides from LWR fuel using molten-salt-based electrochemical processes. *Nuclear Technology* 136:342–353
49. Gourishankar K, Redey L, Williamson M (2002) Electrochemical Reduction of Metal Oxides in Molten Salts. *Light Metals 2002*, TMS
50. Westphal BR, Keiser DD, Rigg RH, Laug DV (1994) Production of Metal Waste Forms from Spent Fuel Treatment. *Proceedings of the DOE Spent Nuclear Fuel Meeting: Challenges and Initiatives*, Salt Lake City, Utah; December 13–16, 1994
51. Abraham DP, McDevitt SM, Park J (1996) Metal waste forms from the electrometallurgical treatment of spent nuclear fuel. *Proceedings of the Embedded Topical Meeting on DOE Spent Nuclear Fuel and Fissile Material Management*, Reno, Nevada, June 16–20, 1996
52. Pereira C, Hash MC, Lewis MA, Richmann MK, Basco J (1999) Incorporation of Radionuclides from the Electrometallurgical Treatment of Spent Fuel into a Ceramic Waste Form. *Materials Research Society Symposium Proceedings* 556 (1999), 115–120
53. Ahluwalia RK, Geyer HK, Pereira C, Ackerman JP (1998) Modeling of a zeolite column for the removal of fission products from molten salt. *Ind Eng Chem Res* 37:145
54. Lexa D, Johnson I (2001) Occlusion and Ion Exchange in the Molten (Lithium Chloride-Potassium Chloride-Alkali Chloride) Salt + Zeolite 4A System with Alkali Metal Chlorides of Sodium, Rubidium, and Cesium. *Metallurgical Mater Trans* 32B:429
55. Phongikaroon S, Simpson MF (May 2006) two site equilibrium model for ion exchange between multivalent cations and Zeolite-A in a molten salt. *AIChE J* 52(5):1736–1743

56. Kim EH, Park GI, Cho YZ, Yang HC (2008) A new approach to minimize pyroprocessing waste salts through a series of fission product removal process. *Nuclear Technol* 162(2):208–218
57. Simpson MF, Sachdev P (2008) Development of electrorefiner waste salt disposal process for the EBR-II spent fuel treatment project. *Nuclear Eng Technol* 40(3), April 2008
58. Simpson MF, Goff KM, Johnson SG, Bateman KJ, Battisti TJ, Toews KL, Frank SM, Moschetti TL, O'Holleran TP (2001) A description of the ceramic waste form production process from the demonstration phase of the electrometallurgical treatment of EBR-II spent fuel. *Nuclear Technol* 134:263–277
59. Thomas JL, Mange M, Eyraud C (1971) *Molecular Sieve Zeolites-I*, R.F. Gould, Ed., Amer Chem Soc (1971), 101:443–449
60. Ebert WE (2005) Testing to evaluate the suitability of waste forms developed for electrometallurgically treated spent sodium-bonded nuclear fuel for disposal in the Yucca Mountain repository. Argonne National Laboratory, ANL-05/43, September 2005

Nuclear Fusion

THOMAS J. DOLAN

Department of Nuclear, Plasma, and Radiological Engineering, University of Illinois, Urbana, IL, USA

Article Outline

Glossary
 Definition of the Subject
 Introduction
 Fusion Reactions
 Magnetic Confinement
 Inertial Confinement Fusion (ICF)
 Challenges of Plasma Theory
 Fusion Technology Issues
 Fusion Reactor Design Studies
 Fusion–Fission Hybrids
 Future Directions
 Acknowledgments
 Bibliography

Glossary

COE Cost of electricity
EAST Experimental Advanced Superconducting Tokamak
ECRH or ECH Electron cyclotron resonance heating – heats plasma electrons at the natural rotation frequency of electrons in a magnetic field

EFDA European Fusion Development Agreement
FRC Field-reversed configuration
ICRH or ICH Ion cyclotron resonance heating – heats plasma ions at the natural rotation frequency of ions in a magnetic field
ICF Inertial confinement fusion
IFMIF International Fusion Materials Irradiation Facility
ITER International Thermonuclear Experimental Reactor
LHCD or LH Lower hybrid resonance wave heating or current drive – Microwaves injected into the plasma induce a plasma current
Magnetic shear Variation of magnetic field direction from one layer to the next – helps to preserve plasma stability
MHD Magnetohydrodynamic model – treats plasma as a conducting fluid
MTF Magnetized target fusion
NBI Neutral beam injection – injection of high-energy neutral atoms into plasma to heat it, and to help control the plasma current and rotation
OH Ohmic heating – resistive heating caused by a current flowing through the plasma
PF Poloidal magnetic field – The magnetic field component that runs the short way around the torus (donut-shaped vessel). See Fig. 6
Q Fusion energy gain ratio or fusion power gain ratio = (fusion energy)/(input energy) or (fusion power)/(input power)
RF, rf Radiofrequency
SC Superconducting magnet coils – Coils that have zero resistance at very low temperature ($T \sim 4$ K), usually made of Nb₃Sn or NbTi in a copper matrix
TF Toroidal magnetic field – The magnetic field component that runs the long way around the torus (donut-shaped vessel). See Fig. 6

Symbols

Symbol	Units	Meaning
a	m	Minor plasma radius at plasma edge (for circular plasma cross section) (Fig. 4)
B	T	Magnetic field
beta, β	None	Ratio of (plasma pressure)/(magnetic field pressure)

D		Deuterium or deuteron
I_p	MA	Maximum plasma current
L	m	Plasma length
m_e	kg	Electron mass
m_i	kg	Ion mass
n	m^{-3}	Plasma electron density (electrons per m^3)
r	m	Minor plasma radius (Fig. 4)
R, R_0	m	Major plasma radius, and its value at the center of the plasma (Fig. 4)
T	C, K, or keV	Temperature. 1 keV = 11.6 MK (Mega-Kelvin)
T		Tritium or triton
T_e	keV	Electron temperature
T_i	keV	Ion temperature
$v_{ }$	m/s	Particle velocity component along B field direction
v_{\perp}	m/s	Particle velocity component perpendicular to B field

Definition of the Subject

Nuclear Fusion of hydrogen isotopes into helium and heavier elements is the energy source of our sun and the other stars. The goal of nuclear fusion research is to build “miniature suns” on earth that will provide clean energy for electricity generation, production of hydrogen for fuel, and industrial applications.

Introduction

The joining together of two light elements to produce a heavier element is called a “nuclear fusion reaction.” For example, a deuteron plus a triton can produce ^4He plus a neutron



where D = deuteron, T = triton, n = neutron, the reaction product kinetic energies are given in parentheses, and 1 MeV = 1.602×10^{-13} J. (Deuterium and tritium are isotopes of hydrogen having one neutron or two neutrons, respectively.)

The positively charged atomic nuclei repel each other, so they must have high speeds in order to

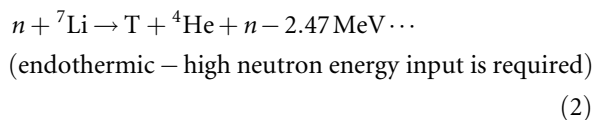
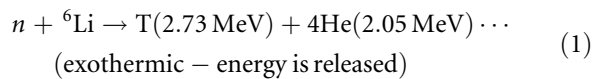
approach close enough for reactions to occur, which correspond to high temperatures. Temperatures are usually expressed in keV, where 1 keV = 11.6 MK (Mega-Kelvin).

One liter of water contains 0.034 g of deuterium. When burned in a nuclear fusion reactor, this deuterium could yield as much energy as burning 300 liters of gasoline. Thus, there is enough deuterium in the rivers, lakes, and oceans of the world to provide the world’s energy needs for millions of years. The first generation of fusion power plants will use deuterium and tritium fuels. Tritium is unavailable on earth, but it can be bred by neutron absorption in lithium, so deuterium and lithium are the primary fuels. Table 1 lists some potential benefits of such fusion reactors.

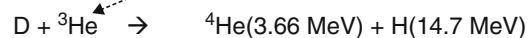
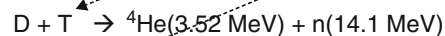
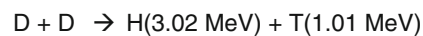
Although the fuel for fusion reactors would be cheap, the reactor capital costs would be very high, so the cost of electricity could still be expensive (to be discussed later).

Fusion Reactions

Tritium has negligible abundance on earth, but it can be bred from lithium via the reactions



The fusion reactions with the highest probabilities at $T < 80$ keV are



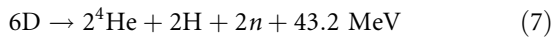
where H = proton, D = deuteron, T = triton (distinguish from T meaning temperature), n = neutron, and the energies of the reaction products are shown in parentheses. The TT reaction has a probability similar to those of the DD reactions. The two branches of the DD reaction have roughly equal probabilities. If the

Nuclear Fusion. Table 1 Potential benefits of fusion reactors

- Abundant fuels (deuterium and lithium) – enough for millions of years
- Cheap fuel
 - D fuel cost $\approx 3 \times 10^{-14}$ \$/J (assuming D at \$10,000/kg)^a
 - Lithium fuel cost $\approx 10^{-12}$ \$/J (assuming Li at \$300/kg)^a
 - Coal cost $\approx 2 \times 10^{-9}$ \$/J (assuming \$150/t)
 - Gasoline cost $\approx 2 \times 10^{-8}$ \$/J (assuming 1.00 \$/l)
- Fuel available to all nations, could reduce conflicts over fossil fuels
- Fuel has low mass, easy to transport
- Clean – avoids pollutant emissions, less waste than fission or coal
- Safe – no supercriticality hazard or meltdown hazard
- Does not require expensive energy storage
- No high-level radioactive waste
- Neutrons from fusion reactors could be used to de-activate high-level radioactive waste from fission reactors or to breed fission reactor fuel

^aAssuming a price of \$10,000/kg of D and an energy yield of 7.2 MeV per deuteron in a catalyzed DD fuel cycle (described below), this corresponds to a fuel cost of 3×10^{-14} \$/J. Early fusion reactors will use lithium to breed tritium, because it greatly enhances the fusion reaction rate. In 2008, the average price of lithium carbonate was about 4–6 \$/kg in the USA, so the cost of the lithium itself was about 20–40 \$/kg [1]. Future demand for lithium batteries might escalate this price as high as 300 \$/kg. Assuming 17.6 MeV released from each tritium bred from lithium, this corresponds to a fuel cost of about 10^{-12} \$/J, as shown in the table

³He and T produced by the DD reactions are recycled as fuel in the other two reactions (dashed arrows), the sum of these four reactions yields



which is called the “catalyzed DD” fuel cycle, having an average energy yield of 7.2 MeV per deuteron. The only fuel input would be deuterium, avoiding the need to breed tritium from lithium, but the fusion power density would be much lower than from the DT fuel cycle using tritium bred from lithium.

Other reactions, such as $\text{H} + {}^6\text{Li}$ and $\text{H} + {}^{11}\text{B}$, have lower reaction probabilities at $T < 100$ keV, so they are more difficult to use, but they would have the potential advantage of emitting fewer (or no) neutrons, which lead to the production of radioactive materials.

Fusion Reactor Requirements

In order to build a fusion reactor two requirements must be met, heating and confinement:

Heating Requirement The fuel must be heated to $T \sim 15$ keV (170 Million Kelvin) for the D + T reaction, and $T \sim 40$ keV (460 MK) for the D + D fuel cycle. Ions with more positive charge, such as lithium and boron, repel each other more strongly, so higher temperatures would be required to use them as fuels. (At such high

temperatures, hydrogen is in the *plasma state* [a sea of positive ions and negative electrons]. The five states of matter are solid, liquid, gas, plasma, and “dark matter.” Other familiar forms of plasmas [some of them only partially ionized gases] include lightning, welding arcs, hot flames, and fluorescent lights.)

Table 2 lists some plasma heating methods.

Confinement Requirement The fuel must be confined long enough for part of it to “burn” (to fuse) before it is lost or cools off.

The heating and confinement requirements for a DT reactor may be expressed in the “triple product” parameter

$$n\tau_E T > 4 \times 10^{21} \text{ m}^{-3} \text{ s keV} \quad (8)$$

where n is fuel ion density (ions/m³), τ_E is their energy confinement time (s), and T is their temperature (keV). (The required product of density and confinement time is often referred to as the “Lawson criterion.”) The energy gain ratio is defined to be

$$Q = \frac{(\text{fusion energy per pulse})}{(\text{input energy per pulse})}$$

or

$$Q = \frac{(\text{fusion power})}{(\text{input power})}$$

if the reactor operates steady state.

Nuclear Fusion. Table 2 Plasma heating methods

Ohmic heating	OH	Ohmic heating power density $P_{\text{oh}} = \eta J^2$ (W/m ³), where J = current density (A/m ²) and η = plasma resistivity (ohm-m). At high temperatures η is very low, and ohmic heating is ineffective.
Neutral beam injection	NBI	Ions accelerated to high energies (typically 0.04–1 MeV) pass through a gas cell where some of them are neutralized. The neutral atoms can then penetrate across the magnetic field into the plasma, where they become ionized and trapped, depositing their energy in the plasma.
Ion cyclotron resonance heating	ICRH	Radio waves are injected into the plasma at approximately the frequency of the ion spiral motion around the magnetic field lines, accelerating their motion. Frequency $f = eB/2\pi m_i$ (Hz), where e = electronic charge, B = magnetic field, m_i = ion mass.
Electron cyclotron resonance heating	ECRH	Microwaves injected into the plasma with near the frequency of electron spiral motion around the magnetic field lines. $f = eB/2\pi m_e$ (Hz), where m_e = electron mass.
Lower hybrid wave heating	LH	Waves are injected at the lower hybrid frequency, which is between the ECRH and ICRH frequencies.
Compression		If the plasma is compressed, such as by increasing the magnetic field or by squeezing a metallic shell around the plasma, its temperature and pressure increase.

For the DT fusion reaction Q is given approximately by the equation [2]

$$Q \approx 5(n\tau_E T) / [5 \times 10^{21} - n\tau_E T] \quad (9)$$

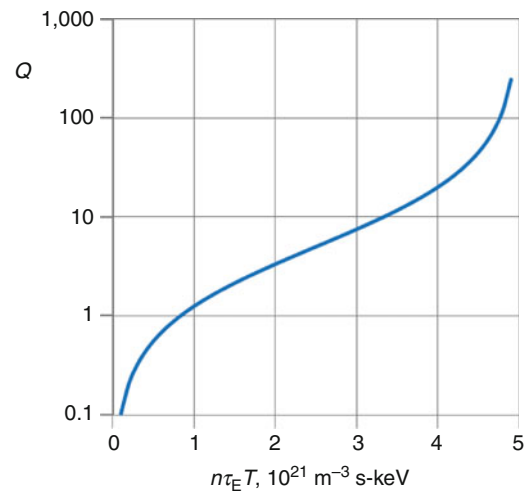
which is illustrated in Fig. 1.

Thus, to attain $Q > 10$, values of $n\tau_E T \sim 4 \times 10^{21} \text{ m}^{-3} \text{ s-keV}$ are needed. (This equation and graph are only approximate, varying with the type of plasma confinement system.)

When confinement is very good, the alpha particles (^4He ions) produced by fusion reactions can deposit enough energy in the plasma to sustain its temperature. Then the input power for heating may be turned off. (Some input power may still be needed to drive a plasma electrical current.) This self-sustainment condition, called “ignition,” corresponds to the right side of Fig. 1, where Q reaches very high values. The precise value of the triple product that produces ignition varies from one type of confinement system to another and with plasma impurity content.

For DD reactions, higher values of $n\tau_E T$ are required, so the DT reaction will probably be used in the first generation fusion power plants.

Everything in nature tends toward a state of *thermodynamic equilibrium*, in which the temperatures of various bodies are uniform and equal to each other. Since we want to keep the chamber wall temperature below about 1,200 K and the fuel ion temperature



Nuclear Fusion. Figure 1
Energy gain ratio Q versus triple product

$>10^8$ K, the plasma confinement system must retard the establishment of thermodynamic equilibrium. Coulomb collisions and plasma instabilities bring the plasma closer to thermodynamic equilibrium, so avoiding it is a difficult problem.

Plasma Confinement Methods

Plasma confinement methods are listed in Table 3.

Nuclear Fusion. Table 3 Plasma confinement methods

Solid walls	Low-temperature plasmas, such as fluorescent lights, may be contained by glass or metal tubes. Hot plasma confinement in magnetic fields may be augmented by solid walls for brief periods of time, but prolonged contact cools the plasma rapidly by heat conduction and may overheat the wall.
Gravity	Although stellar plasmas, such as the sun, are confined by gravity, the mass of a laboratory plasma is far too small for self-gravitational attraction to be significant. (The mass density and temperature at the center of the sun are about 150 g/cm^3 and 1.3 keV).
Inertia	If a DT fuel pellet is quickly compressed to ultrahigh densities, significant fusion burn can occur before the compressed pellet expands. The inertia of the pellet limits the expansion rate of the internal plasma. At $n = 10^{30} \text{ m}^{-3}$, and $T = 10 \text{ keV}$, a confinement time $\tau_E \sim 4 \times 10^{-10} \text{ s}$ would yield a triple product $n\tau_E T \sim 4 \times 10^{21} \text{ m}^{-3} \text{ s-keV}$. The compression may be produced by laser beams or by particle beams.
Radiofrequency waves	Radiofrequency waves and microwaves can confine low-pressure plasma, but enormous power inputs would be required to confine high-pressure plasmas. These waves can augment magnetic confinement.
Electrostatic fields	Electrostatic potential peaks can be established with high-voltage grids or by creating local regions of higher plasma density. Positive peaks can inhibit ion motion, and negative peaks can inhibit electron motion. Purely electrostatic plasma confinement has only succeeded in confining low-pressure plasmas, but such electrostatic potential peaks can be used to augment magnetic confinement.
Magnetic fields	Confinement by strong magnetic fields is the most promising means for prolonged containment of high-pressure plasmas. It is based on the fact that ions and electrons cannot travel across a magnetic field easily. Instead, they tend to travel in helical paths along the field lines.

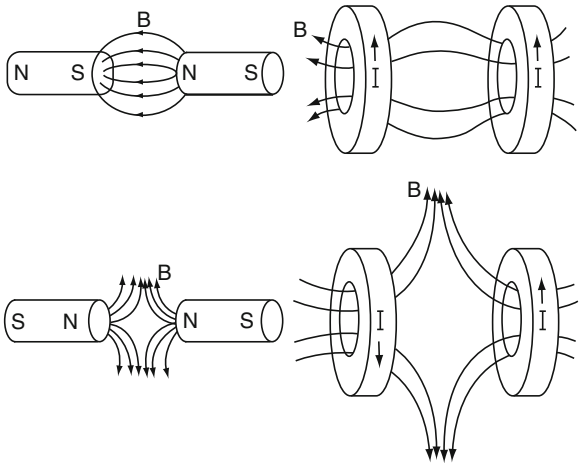
Magnetic Confinement

Magnetic fields may be either “open” (Figs. 2 and 3) or “closed” (Fig. 4).

A plasma ion starting out at the center with velocity components $v_{\parallel 0}$ parallel to magnetic field and $v_{\perp 0}$ perpendicular to the magnetic field would experience a higher magnetic field as it moves toward the magnet coil. In the higher field, its rotational velocity component v_{\perp} would increase, and its parallel velocity component v_{\parallel} could decrease gradually to zero, where it would be reflected back toward the center (hence, the name “magnetic mirror”). The ion would oscillate back and forth between points a and b, restrained by a magnetic field gradient force

$$F = - (m_i v_{\perp}^2 / 2B) \nabla B \quad (10)$$

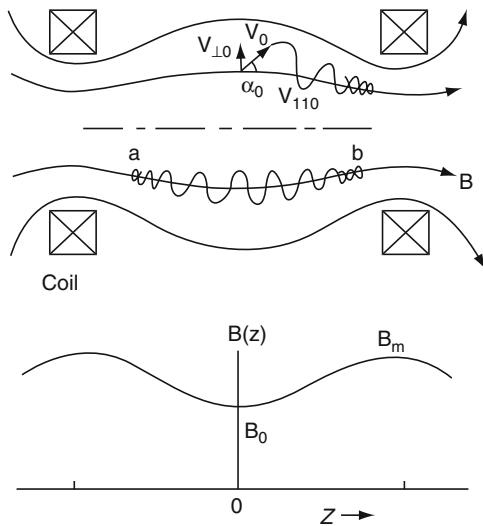
Electrons would also be confined in the same way. Although electrons and ions are reflected by high magnetic fields, those with high velocities $v_{\parallel 0}$ along the field



Nuclear Fusion. Figure 2 Simple magnetic “mirror” fields **B** (top) and “spindle cusp” fields (bottom) produced by bar magnets (left side) or by a pair of circular magnet coils carrying currents **I** (right side). Plasma could be confined in the central regions

lines will escape, and confinement is limited by the time it takes for Coulomb collisions to increase their parallel velocities. The end loss problem can be eliminated by using a closed magnetic field, Fig. 4.

It might appear that electrons and ions could be confined forever as they spiral along the closed magnetic field lines, but the magnetic field gradient and curvature cause a drift velocity across the magnetic field. This drift can be compensated by twisting the magnetic field lines, producing a helical closed magnetic field, Fig. 5.



Nuclear Fusion. Figure 3

A simple magnetic mirror (top) and axial variation of magnetic field strength (bottom)

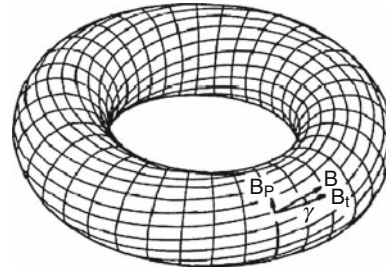
This twisting of the magnetic field, called “rotational transform,” may be produced by inducing a plasma electrical current in the toroidal direction (a “tokamak”), or by using specially shaped magnetic field coils (a “stellarator”).

There are many processes that cause plasma energy loss, Table 4.

In what follows, we describe tokamaks, stellarators, compact toroids, and open magnetic confinement systems, then inertial confinement systems, other fusion concepts, plasma theory issues, fusion technology issues, and fusion reactor design studies.

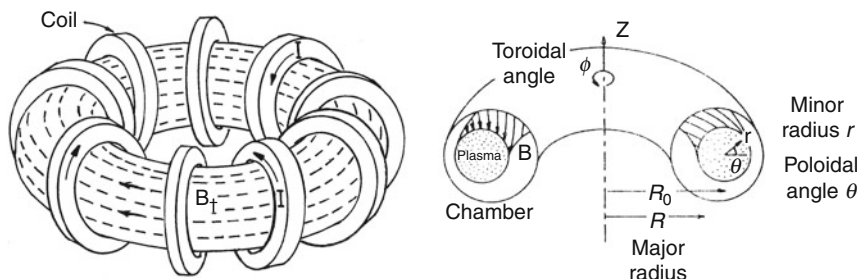
Tokamaks

The name “tokamak” is an acronym for the Russian “тороидальная камера с магнитными катушками”



Nuclear Fusion. Figure 5

A helical magnetic field. The field lines B have components B_t in the “toroidal” direction (the long way around the torus) and B_p in the “poloidal” direction (the short way around the torus). Twisting the magnetic field compensates for the tendency of particles to drift across the magnetic field



Nuclear Fusion. Figure 4

“Closed” (toroidal) magnetic field lines B_t (dashed lines) generated by circular magnet coils carrying currents I , and the definitions of “major radius R ” and “minor radius r .” The value of r at the plasma edge is called “ a ,” and the value of R at the plasma center is called “ R_0 ”

Nuclear Fusion. Table 4 Plasma energy loss mechanisms

1. Particle and energy loss along magnetic field lines (open magnetic systems)
In simple magnetic mirrors, the loss time is roughly the time for collisions to increase the ion velocity along the field lines. If electrostatic potential barriers are used, then the loss time is roughly the time for collisions to increase the particle energies over the barriers.
2. Particle drift velocities across the magnetic field, driven by magnetic field gradient and curvature and by an electric field.
3. Magnetohydrodynamic (MHD) plasma instabilities, whereby the plasma boundary, affected by gradients of pressure, current density, and magnetic field, pushes through the magnetic field toward the chamber wall.
4. Radiation losses from line radiation, bremsstrahlung radiation (radiation from electrons during deceleration by collisions), cyclotron radiation, and recombination radiation. Radiation losses can be high when the plasma contains many impurity ions (such as iron), when the temperature is very high, or when the magnetic field is very strong (cyclotron radiation).
5. Energy transport across the magnetic field by heat conduction and convection.
6. Interaction of electrons and ions with plasma waves can enhance energy transport. Such “microinstabilities” have required many decades of plasma theory, experiment, and computer simulation to be understood.
7. Ion energy loss by charge exchange. (A hot ion grabs an electron from an atom, becomes neutralized, and is no longer confined by the magnetic field, which does not confine neutral atoms.)

(toroidal'naya kamera s magnitnymi katushkami – toroidal chamber with magnetic coils). Figure 6 shows the main components of a tokamak.

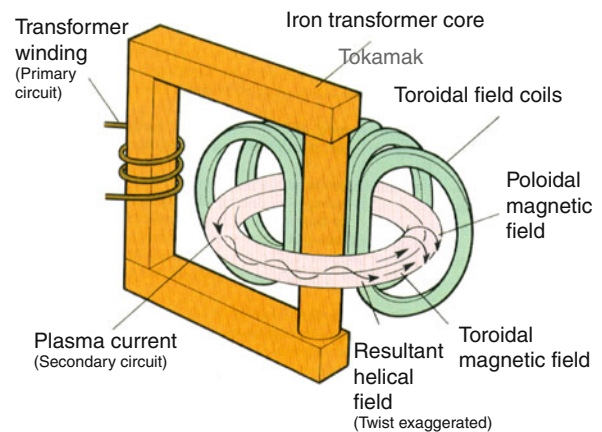
Pulsing a current in the primary winding of the transformer induces a high current in the plasma, which generates the poloidal magnetic field. (Although this figure shows an iron core transformer, most modern tokamaks use air core transformers.) When the transformer saturates, the current gradually dies away. A “current drive” system, such as electromagnetic waves or neutral beam injection, is needed to sustain the current for long periods. Although Fig. 6 shows a plasma with circular cross section, most modern tokamaks have noncircular cross sections, in order to facilitate higher beta values (ratios of plasma pressure to magnetic field pressure).

Over 200 tokamaks have been built worldwide. Table 5 shows the parameters of a few of them.

Figure 7 shows the interior of the European Fusion Development Agreement (EFDA) Joint European Torus (JET), which has a noncircular cross section.

The walls are lined with refractory tiles to withstand high heat fluxes. Using DT fuel JET briefly generated 16 MW of heat from fusion reactions. JET has also demonstrated the feasibility of using beryllium tiles on the walls.

The National Spherical Torus Experiment (NSTX) and Mega Ampere Spherical Torus (MAST) are smaller,

**Nuclear Fusion. Figure 6**

Simplified diagram of a tokamak with circular plasma cross section (Courtesy of Culham Centre for Fusion Energy, UK)

with very low aspect ratios $R_0/a \sim 1.4$, which facilitates high values of β (the ratio of plasma pressure to magnetic field pressure). They are called “spherical tokamaks.”

In order to attain higher triple product values, larger experiments are needed. The International Thermonuclear Experimental Reactor (ITER) Joint Project began construction in 2008 with an estimated cost of about 5 billion Euro. It is to begin operation in about

Nuclear Fusion. Table 5 Parameters of some tokamaks. SC means superconducting coils. Parameters of future experiments are planned values [3]

Name	Country	Years	R_0/a , m/m	B , T	I_p , MA	ECRH + ICRH + LH, MW	NBI, MW	Remarks
TFTR	USA	1982–1997	2.4/0.8	6	3	11	39	Circular cross section plasma; achieved 10 MW fusion power for fraction of a second
DIII-D	USA	1986–	1.66/0.67	2.2	3	11	20	Beta = 12%, graphite walls, plasma shaping
Tore Supra	France	1988–	2.25/0.7	4.5 SC	2	16	1.7	Pulses lasting many minutes
T-15	Russia	1988–1995, 2009–	2.43/0.42	3.5 SC	1	11	9	5 s pulses
JT 60U	Japan	1991–2010	3.4/1.0	4.2	5	24	50	Record fusion triple product, would yield $Q = 1.2$ if DT fuel were used
JET	UK	1992–	3/1.25	4	6	29	24	Generated 22 MJ of DT fusion energy in 4 s. Has used Be wall tiles
MAST	UK	1999–	0.85/0.6	0.5	1.4	1	4	Spherical tokamak
NSTX	USA	1999–	0.85/0.6	0.6	1.4	3	7	Spherical tokamak
EAST	China	2006–	1.75/0.43	5 SC	0.5	8		First fully superconducting tokamak with noncircular cross section
KSTAR	Korea	2008–	1.8/0.5	3.5 SC	2	13	14	Fully superconducting coils
SST-1	India	2012–	1.1/0.2	3 SC	0.2	3	0.8	1,000 s pulses
JT 60 SA	Japan	2014–	3.16/1.02	2.7 SC	5.5	7	34	100 s pulses, startup 2014
ITER	France	2019–	6.2/2.0	5.3 SC	15	40	33	Startup 2019. 400 MW, 400 s pulses in 2023

2019, reaching full power operation in about 2027. Figure 8 shows how the experimentally attained values of the triple product have increased over the years of fusion research.

ITER should demonstrate a power gain ratio $Q > 10$ for hundreds of seconds, and $Q > 5$ for longer periods. The next step planned after ITER will be a technology demonstration fusion power plant called “DEMO” that generates electricity for commercial use.

Stellarators

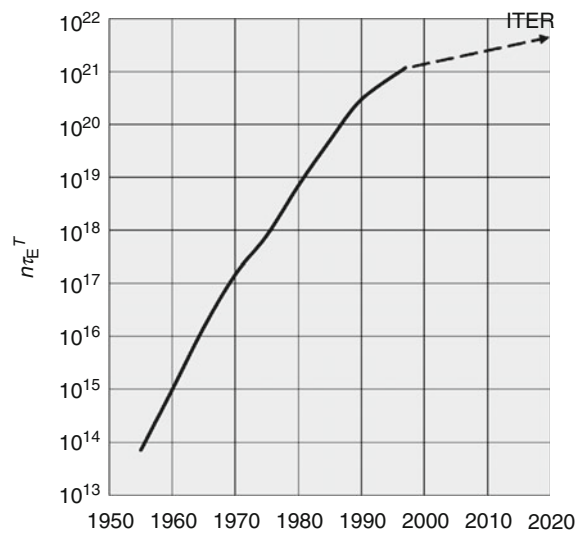
Stellarators (including “heliotrons” and “torsatrons”) produce rotational transform by shaping the magnet coils, Figs. 9 and 10.

The Large Helical Device at the National Institute for Fusion Sciences in Japan, Fig. 11, is a “heliotron” (torsatron) with two helical coils.

Figure 12 shows the Wendelstein 7-X experiment, which uses modular coils to achieve a magnetic field shape suitable for plasma confinement.

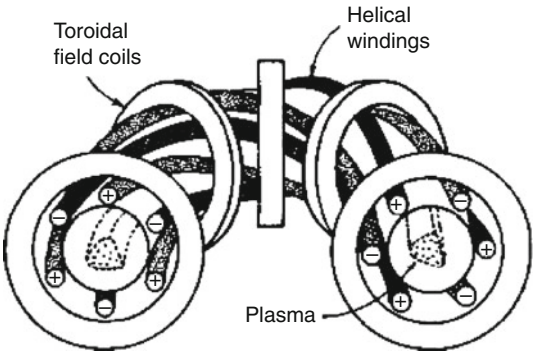


Nuclear Fusion. Figure 7
The interior of the Joint European Torus (JET) (Courtesy of Culham Centre for Fusion Energy, UK)

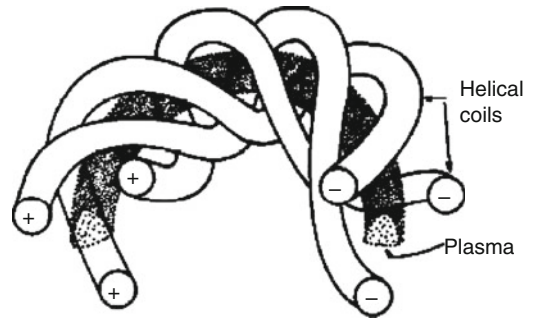


Nuclear Fusion. Figure 8
Triple product $n\tau_E T$ (m⁻³ s-keV) attained by tokamak experiments versus year. The dashed arrow points to planned ITER values [4]

Table 6 shows some parameters of these stellarators. Since stellarators do not require a strong plasma current, they have the following advantages over tokamaks:



Nuclear Fusion. Figure 9
Magnet coils of a stellarator with three pairs of helical windings, which have currents in alternating directions (indicated by + and - signs). Other numbers of helical windings are also possible



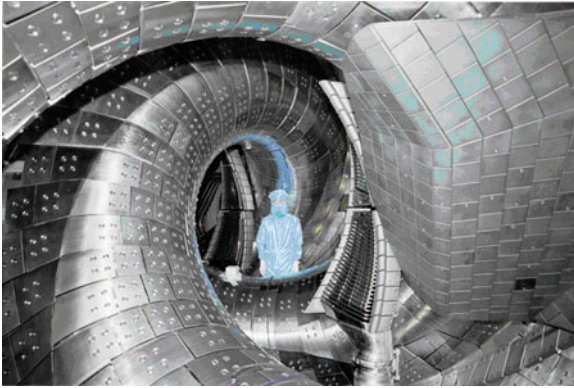
Nuclear Fusion. Figure 10
A torsatron with three helical coils and no toroidal field coils

- No strong disruptions of the plasma.
- Current drive is not required, so the input power is lower, and the energy gain ratio Q can be higher.
- Potentially less plasma turbulence and better energy confinement.

One the other hand, the coils of stellarators are difficult to wind, join, support, and align, and space for energy recovery blankets is limited.

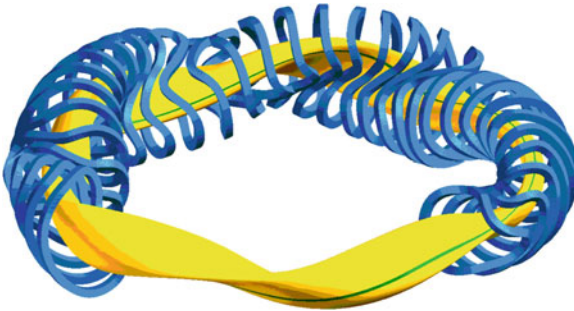
Reversed Field Pinch (RFP)

A reversed field pinch is a toroidal plasma with a relatively weak toroidal field and a strong plasma current. The plasma currents evolve into a stable



Nuclear Fusion. Figure 11

The two helical superconducting coils inside the Large Helical Device, operating since 1998 (Courtesy of the National Institute for Fusion Science, Japan)



Nuclear Fusion. Figure 12

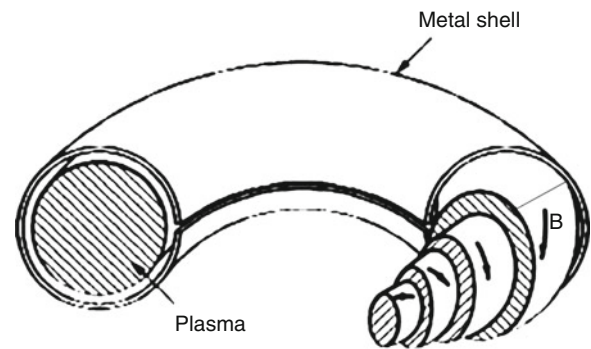
The coils (blue) of the Wendelstein 7-X experiment in Greifswald, Germany, and the resulting plasma shape (orange), expected to begin operation in 2014 (Courtesy of the Max-Planck-Institut fuer Plasmaforschung, Greifswald, Germany)

configuration in which the direction of the toroidal component of the magnetic field at the plasma edge is in the opposite direction from the internal toroidal field, Fig. 13.

This strong variation of magnetic field direction with radius, called magnetic shear, helps preserve plasma stability. This type of experiment has been operated in Italy, Japan, the UK, the USA, and elsewhere. It offers the potential of a compact reactor with a simpler toroidal field system than tokamaks, but plasma energy confinement times may be shorter, and

Nuclear Fusion. Table 6 Main parameters of the LHD and Wendelstein 7-X experiments. The LHD values shown here were not all attained simultaneously. The Wendelstein 7-X parameters are planned values

Parameter	Units	LHD	Wendelstein 7-X
Startup		1998	2014
Major radius R	m	3.7	5.5
Approximate minor radius a	m	0.64	0.53
Magnetic field on axis	T	2.8	3
Heating power	MW	20	15
Average plasma density n	m^{-3}	4×10^{19}	$\sim 10^{20}$
Ion temperature	keV	13.6	(Several)
Electron temperature	keV	10	(Several)
Energy confinement time	s	0.36	0.15
Beta	%	5	



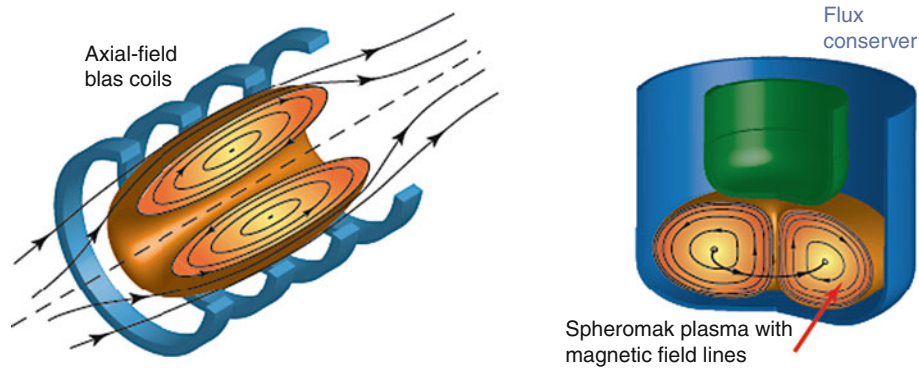
Nuclear Fusion. Figure 13

A reversed field pinch, showing how the direction of the magnetic field varies with radius [5]

efficient current drive to sustain the RFP is difficult to achieve.

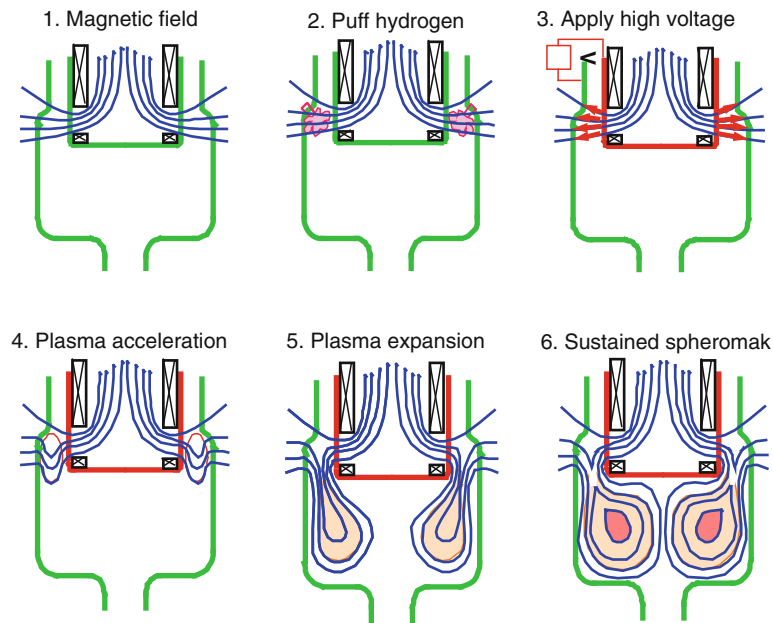
Compact Toroids

Figure 14 compares a field-reversed configuration (FRC) and a spheromak.



Nuclear Fusion. Figure 14

A field-reversed configuration (*left*) and a spheromak. Neither has any structure on the axis of the torus [6]



Nuclear Fusion. Figure 15

Formation of a spheromak by coaxial plasma gun injection into a flux conserving chamber (Courtesy of Lawrence Livermore National Laboratory)

The FRC has only poloidal magnetic field components, with zero (or nearly zero) toroidal field. The spheromak is a naturally stable configuration having both toroidal and poloidal magnetic field components. [Figure 15](#) illustrates one way of generating a spheromak, by using a coaxial plasma gun.

In step (1). A coil (black) generates a magnetic field (blue). (2) A puff of hydrogen is admitted into the

vacuum (pink cloud). (3) A high voltage is applied between the inner and outer cylinders, ionizing the gas. (4) The plasma current flowing interacts with the magnetic field to produce a thrust that accelerates the plasma downward. (5) The plasma (pink) expands into the metal chamber (green). (6) Image currents in the walls create an opposing magnetic field that retards the plasma penetration into the wall. This is

called a “sustained spheromak” configuration with quasi-closed field lines similar to tokamaks. The plasma has both toroidal and poloidal magnetic fields without the need for toroidal field coils. It is therefore simpler than tokamaks, and potentially less expensive. Some parameters of the Sustained Spheromak Physics Experiment (SSPX) are shown in Table 7.

Spheromak experiments generally last only a few milliseconds, with electron temperatures less than

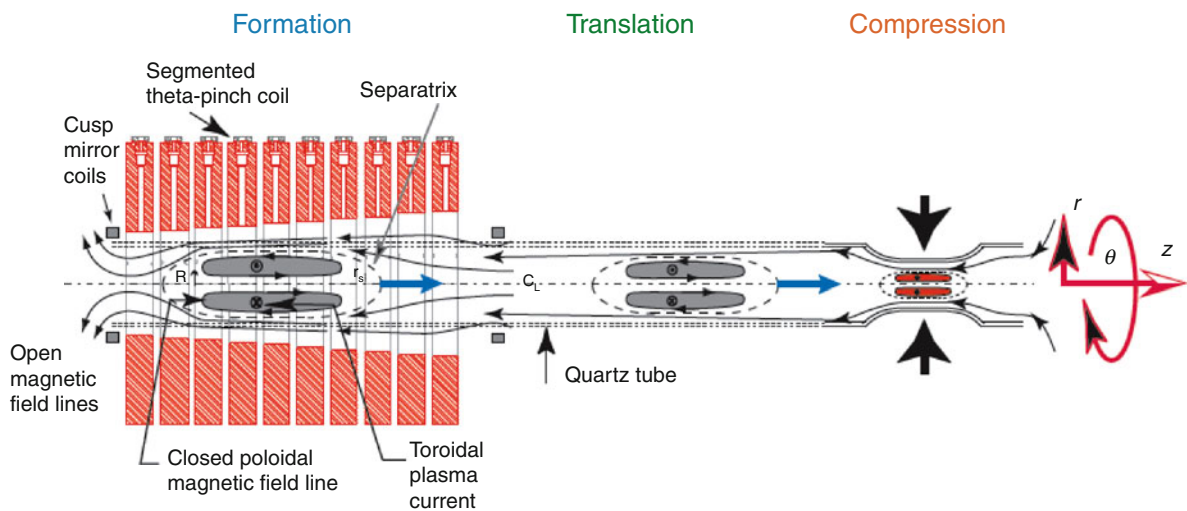
Nuclear Fusion. Table 7 Typical SSPX parameters (SSPX is now decommissioned) (Courtesy of Lawrence Livermore National Laboratory)

Flux conserver radius and height	0.5 m × 0.5 m
Radius of magnetic axis	0.31 m
Plasma minor radius	0.17 m
Peak discharge current	0.45 MA
Toroidal current	0.6 MA
Peak toroidal field	0.6 T
Edge poloidal field	0.35 T
Plasma duration	4.5 ms
Plasma density	$5 \times 10^{19} \text{ m}^{-3}$
Peak electron temperature	0.35 keV

1 keV. Larger spheromak experiments with more heating auxiliary power and longer duration would be needed to test this concept adequately.

Field-Reversed Configuration (FRC) A field-reversed configuration may be formed by a theta pinch coil, as shown in Fig. 16.

A theta pinch coil is usually a one-turn coil with a sudden current flowing in the theta (azimuthal) direction. The coil current may rise to its peak value in microseconds, squeezing the plasma, hence the name “pinch.” With careful programming of the coil current, the plasma may form into an elongated torus with the internal magnetic field in the opposite direction from the external magnetic field (indicated by the small arrows in Fig. 16), hence the name “field reversed.” FRC plasmas can also be generated and sustained by rotating magnetic fields. They appear to be stable, in spite of their weak or zero toroidal magnetic fields. After formation, an FRC plasma may be moved axially to another chamber, where it can be compressed by a converging magnetic field or by an imploding metallic liner [7]. FRC experiments have been conducted at Los Alamos National Laboratory, University of Tokyo, Osaka University, University of Washington, and TRINITY Laboratory (Troitsk, Russia).



Nuclear Fusion. Figure 16

Formation of an FRC by a theta pinch coil, translation to another chamber, and compression by a metallic liner (Courtesy of Los Alamos National Laboratory)

Magnetized Target Fusion (MTF) MTF target plasma parameters could be $n \sim 10^{23} \text{ m}^{-3}$, $T \sim 0.3 \text{ keV}$, $B \sim 3 \text{ T}$, and duration $t \sim 10 \text{ } \mu\text{s}$ before compression. The target plasma could be moved axially into another chamber (Fig. 16), where a metallic liner would compress it to thermonuclear temperature ($T \sim 10 \text{ keV}$). Intense DT fusion burn could occur until the configuration disassembled, and it might be possible to achieve energy gain ratios $Q > 5$ in a less expensive device than tokamaks or stellarators.

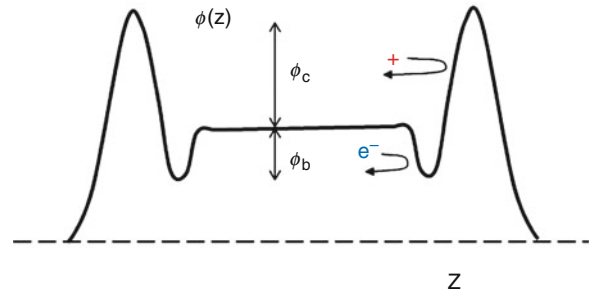
Pulsed, High-Density Fusion A similar concept would accelerate an FRC axially to very high velocity, then inject it into a converging magnetic field, which would compress the plasma up to thermonuclear burn conditions. It would flow through the burn chamber, then continue on into an expansion and exhaust chamber [8].

Open Magnetic Confinement Systems

Open magnetic confinement systems can avoid MHD instabilities if the magnetic field pressure is lower in the plasma confinement region than around the outside, a so-called “minimum-B” confinement system.

Tandem Mirror Due to rapid particle loss along the magnetic field lines, a simple magnetic mirror can only achieve a power gain ratio $Q \sim 1$, which is inadequate for a power plant. This loss can be reduced by creating an electrostatic potential variation along the magnetic field with hills and valleys, as illustrated in Fig. 17.

The positive peaks ϕ_c confine ions in the central cell, and the negative wells ϕ_b confine electrons. The plasma density, temperature, and electrostatic potential in tandem mirrors can be controlled by injection of neutral atom beams, by electron cyclotron resonance heating (ECRH), and by radiofrequency (RF) wave heating at selected axial locations. Such electrostatic potential control could make it possible to reduce the required length of a power plant with $Q \sim 10$ from kilometers to less than 200 m. There have been tandem mirror experiments in Japan, Russia, Korea, and the USA. The Gamma-10 experiment in Japan with length $L = 6 \text{ m}$, $a = 0.36 \text{ m}$, and central cell $B = 0.41 \text{ T}$ has confined plasma with $n \sim 10^{18} \text{ m}^{-3}$, $T_e \sim 0.3 \text{ keV}$, $T_i \sim 1 \text{ keV}$. It has achieved values of $\phi_c > 2 \text{ kV}$ [9].



Nuclear Fusion. Figure 17

Electrostatic potential versus axial position in a tandem mirror

Electrostatically Plugged Cusps High-voltage electrodes can also produce electrostatic potential hills and valleys along magnetic field lines, similar to those of a tandem mirror. The electrons are confined by the magnetic field and the negative electrodes, and the ions are confined in a negative electrostatic potential well created by the electrons' negative charge. For example, the Jupiter-IIM electrostatically plugged magnetic cusp experiment in the Ukraine demonstrated low diffusion rates of plasma across the magnetic field [10].

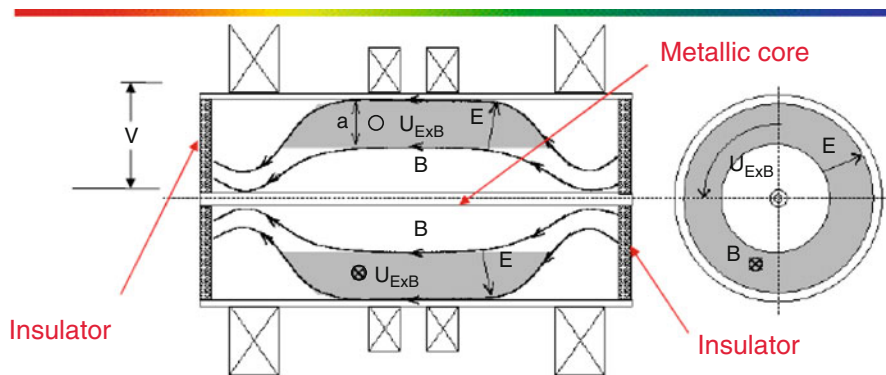
Rotating Plasmas Rotating plasmas can be formed by putting high-voltage electrodes in the ends of a magnetic mirror to create a strong radial electric field, which causes the plasma to rotate azimuthally at high speed. This rotation can accelerate ions up to keV energies, and the angular momentum reduces end losses from the magnetic mirror. Figure 18 shows the Maryland Centrifugal Experiment

This experiment has a central $B \sim 0.3 \text{ T}$, $a \sim 0.2 \text{ m}$, $L \sim 1.4 \text{ m}$, $n \sim 3 \times 10^{20} \text{ m}^{-3}$, $T \sim 20\text{--}60 \text{ eV}$, applied voltage $5\text{--}20 \text{ kV}$, pulse length $1\text{--}10 \text{ ms}$, and has achieved rotational speeds of 100 km/s .

Other Concepts

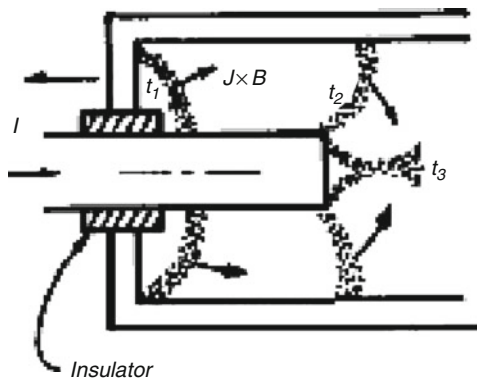
Plasma Focus The plasma focus uses a coaxial plasma gun to accelerate a blob of plasma to high velocities, as illustrated in Fig. 19.

At time t_1 , high voltage breakdown along the insulator forms a plasma that is accelerated axially by the $J \times B$ force. At t_2 the plasma reaches the end of the gun, and at t_3 electromagnetic forces cause it to collapse into a high-density blob, causing a brief burst of fusion



Nuclear Fusion. Figure 18

The Maryland centrifugal experiment (Courtesy of the University of Maryland)



Nuclear Fusion. Figure 19

Plasma acceleration by a coaxial gun

reactions before the plasma expands and cools. At low currents (<0.3 MA), the fusion yield increases proportional to I^4 , where I is the plasma current, but this favorable scaling saturates at high currents, and economical power production appears to be unlikely at reasonably attainable currents (~ 10 MA). Plasma focus devices are useful as sources of neutrons and x-rays, and many have been built around the world.

Levitated Dipole Experiment The Levitated Dipole Experiment magnetically levitates a superconducting coil ring, Fig. 20.

The resulting magnetic surfaces are good for plasma confinement. This experiment has a low-density background plasma plus a hot electron plasma produced by absorption of microwaves. The superconducting ring gradually absorbs heat, so it must be periodically shut

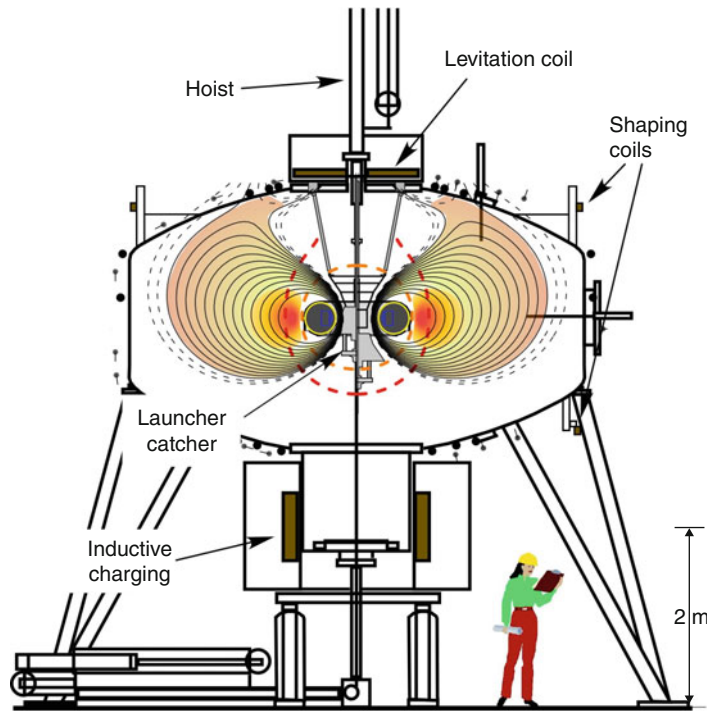
down and recooled. If this scheme were used for a fusion reactor, heating and radiation damage to the ring would be significant issues.

Inertial Electrostatic Confinement (IEC) Plasma may also be confined by concentric spherical grids, Fig. 21.

With tens of kilovolts applied and deuterium plasma, small devices of this type have produced DD fusion reaction neutron emission rates $\sim 10^7$ neutrons/s. In a high-power fusion reactor, however, the electrodes would tend to melt. IEC devices are useful as sources of neutrons and energetic charged particles.

RF Plasma Confinement Potential wells in standing radiofrequency (rf) fields can contain low-pressure plasmas effectively, but for high-pressure plasmas of a fusion reactor excessively high rf power levels would be required. Rf fields have been used to reduce end losses from open magnetic confinement systems, as in the RFC-XX experiment in Japan. High-voltage rf electrodes accelerated ions to higher rotational energies, which inhibited their ability to escape through the point cusps and line cusps.

Polywell The Polywell concept has a spherical array of six or more point cusps, with electron injection creating a negative potential well that focuses ions into the center of the sphere. It is a form of electrostatic plugging of a cusped magnetic field. Good ion focusing to small radii would be required to achieve satisfactory fusion power density.



Nuclear Fusion. Figure 20

The Levitated Dipole Experiment, showing the magnetic surfaces surrounding the levitated coil. The dotted curves indicate the resonant surfaces for ECRH with frequencies 2.45 and 6.4 GHz (Courtesy of Columbia University and Massachusetts Institute of Technology [11])

Muon-Catalyzed Fusion Negative mu mesons (muons) produced by accelerators can bind to deuterons to form atoms. Since the muon is 207 times heavier than an electron, the radius of its orbit is 207 times smaller. This small atomic size allows the shrunken deuterium atom to approach very close to another deuteron or triton, close enough for a nuclear fusion reaction to occur. After a fusion reaction occurs the muon is released, and it can bind to another deuteron (or triton) and catalyze a second fusion reaction. The muon may also be captured by other nuclei, such as by an alpha particle (^4He nucleus), which stops the reaction chain. This “helium sticking” problem has limited the number of experimentally attained fusion reactions to about 100 per muon, which is not enough to achieve an energy gain, since about 6 GeV energy are required to generate each muon by an accelerator, and the DT fusion reaction yields only 17.6 MeV.

Cold Fusion Evidence for low-energy nuclear reactions (LENR) has been reported from many types of

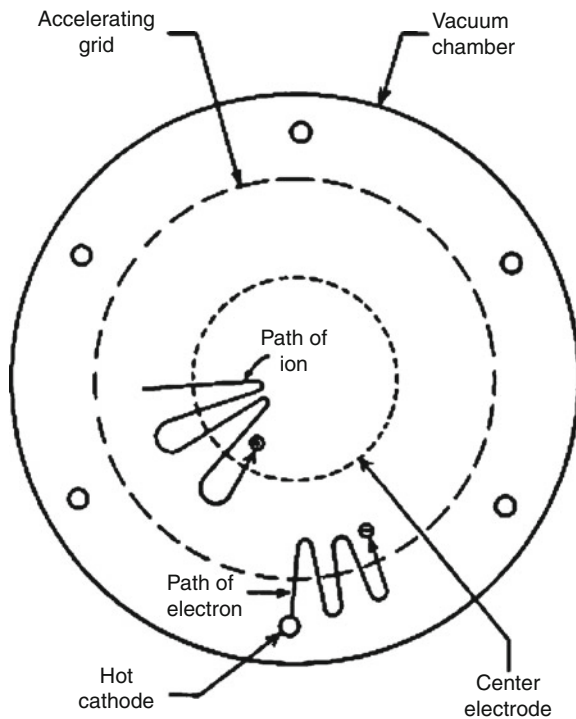
experiments: electrolysis of water, gas discharges, gas diffusion through thin films, electron beam impact, exploding wires, and laser irradiation. The evidence includes apparent heat generation, x-rays, and transmutations. Such phenomena were once called “cold fusion” after the 1989 hypothesis by Fleischmann and Pons that the energy generation by their electrolysis cell was a result of nuclear fusion reactions [12]. Many theories have been proposed to explain the LENR phenomena, but so far, no theory has gained widespread acceptance.

Inertial Confinement Fusion (ICF)

Target Compression

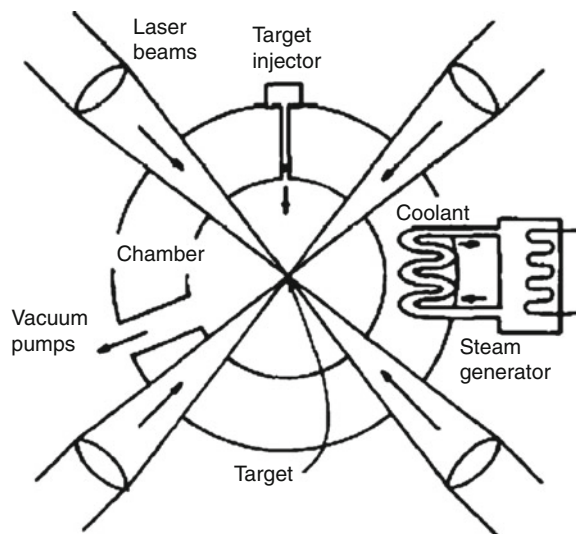
Figure 22 shows the general idea of an inertial confinement fusion (ICF) power plant.

The goal is to ignite frequent small explosions in a blast chamber. The target injector shoots fuel capsules (“targets”) into the chamber several times per second.



Nuclear Fusion. Figure 21

Plasma confinement by spherical grids. In this configuration, the accelerating grid is positive, and the center electrode is negative



Nuclear Fusion. Figure 22

The general idea of a laser ICF power plant

When the target reaches the center of the chamber laser beams ablate its surface, which causes an inwards force that compresses the target to ultrahigh density. Then a high-power laser pulse raises part of the compressed target to ignition temperature, and a thermonuclear burn wave causes a miniature explosion. The explosion energy is deposited in the chamber walls and absorbed in a blanket and coolant, which could be liquid metal, molten salt, or helium gas. The hot coolant boils water in a steam generator, and the steam drives a turbine to generate electricity. (Helium gas coolant could drive a gas turbine directly without needing to make steam.) Instead of using laser beams, the target compression could also be done with heavy ion beams or exploding wires and x-rays. Figure 23 shows the steps of target compression and burn.

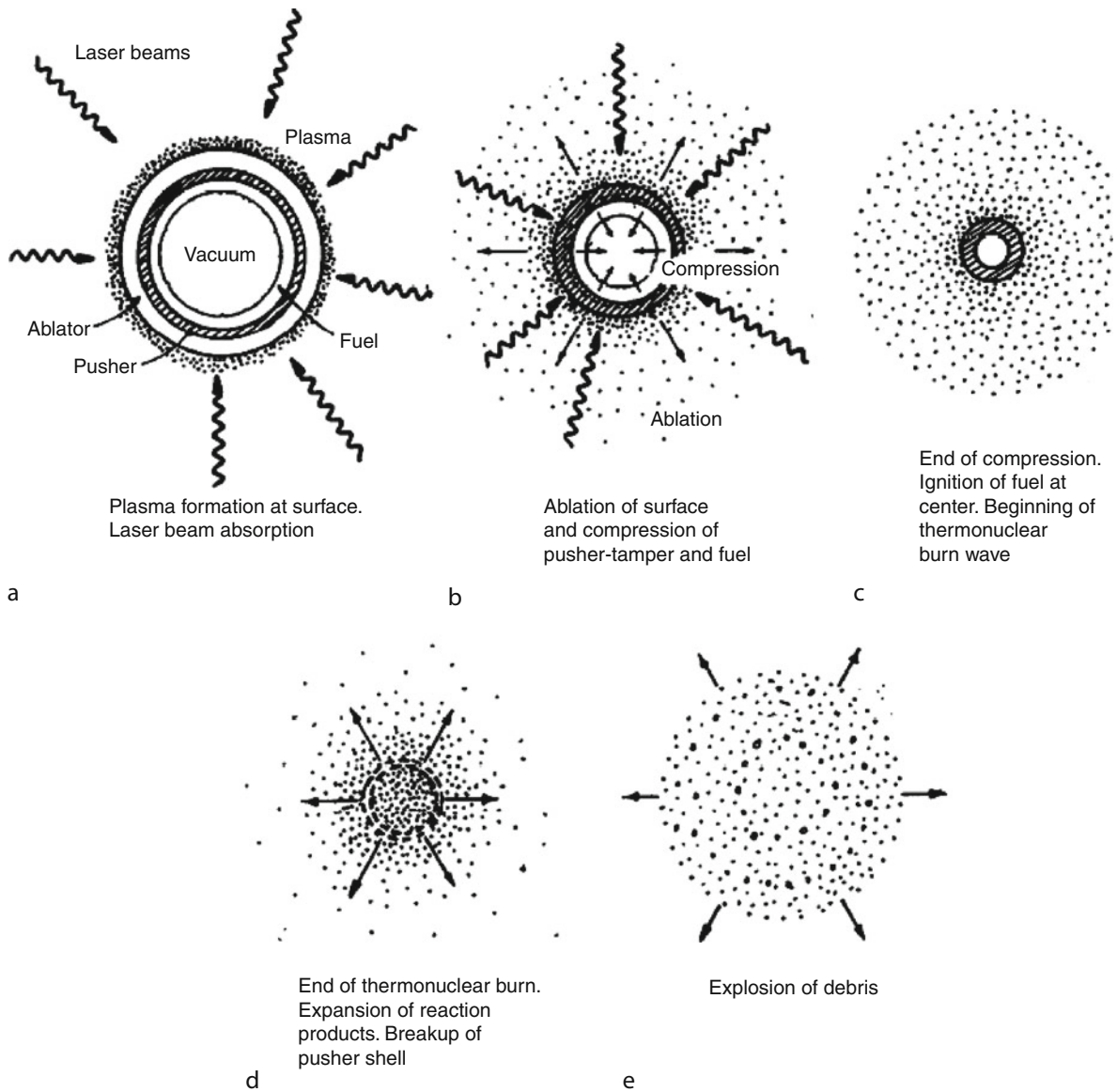
The laser beam heats the plasma surrounding the target shell (Step a of Fig. 23). The plasma ablates the target surface away, causing an inward pressure similar to the forward thrust of a rocket caused by rearward expulsion of burning fuel (Step b). The heating to ignition (Step c) may be done by a shock wave to the center of the volume (Step c) or by a sudden PetaWatt laser beam impact on the outside (“fast ignition”), which requires less heating energy. Then a thermonuclear burn wave can spread to surrounding fuel (Step d).

The National Ignition Facility at Lawrence Livermore National Laboratory, California; the Laser MegaJoule facility in France; and the GEKKO-XII in Japan are all large laser facilities approaching ignition of targets. Figure 24 shows the National Ignition Facility (NIF) in Livermore, California.

Figure 25 shows a man inside the spherical NIF target chamber.

The NIF experiment is expected to demonstrate target ignition and energy gain $Q > 5$ in 2010–2011. It uses glass lasers with low efficiency, and it requires hours between shots to replace the target, which is suspended on a thin quartz fiber. For “ignition campaign” experiments, the target holder must also provide cryogenic refrigeration. The repetition rate is limited by replacement of the target, by evacuation of the chamber, and by cooling of the walls.

If the target capsule irradiation is not extremely uniform, a Rayleigh-Taylor instability [13] can break up the shell before compression is complete, as in Fig. 26.



Nuclear Fusion. Figure 23

Ablation, compression, and burn of an ICF target

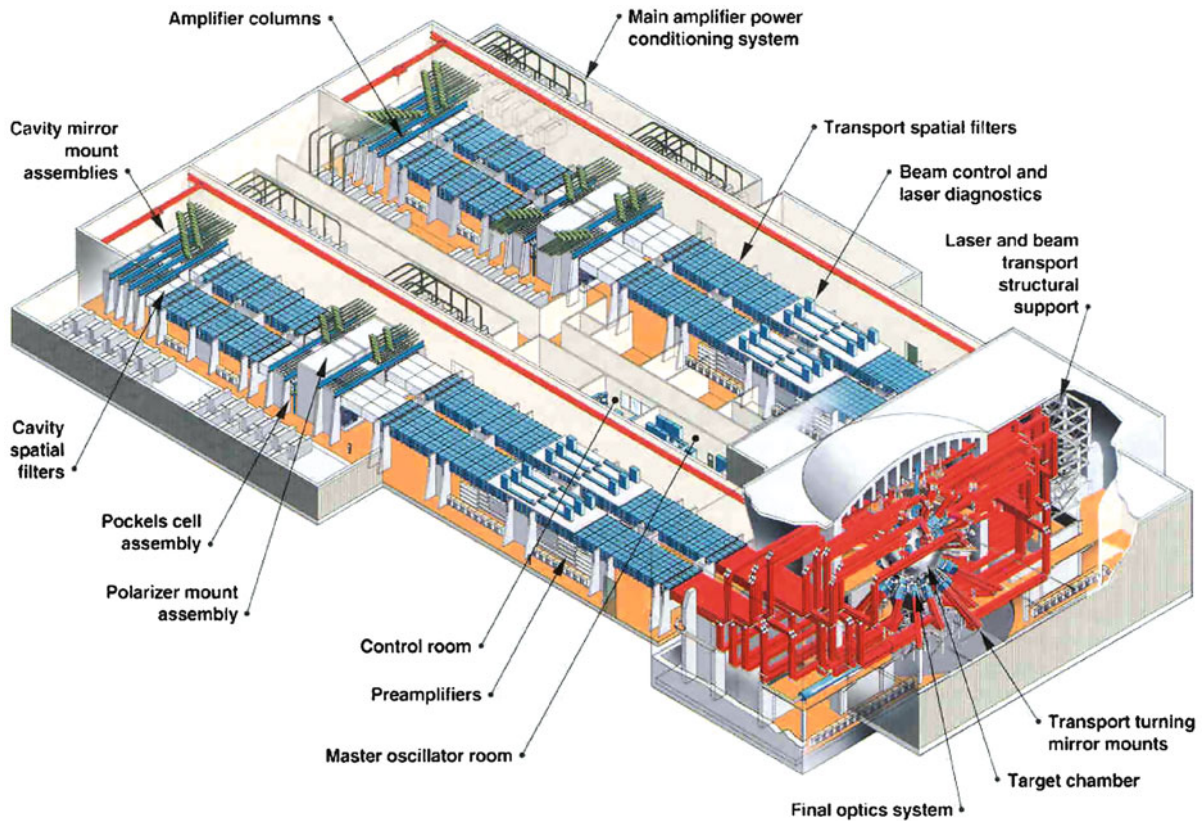
Target pellets may be irradiated directly by the laser beams, or they may be mounted inside a hollow cylinder, called a “Hohlraum,” Fig. 27, to achieve more uniform compression.

The laser beams would enter through holes in the ends of the cylinder and irradiate the inside walls of the Hohlraum, causing them to emit intense x-rays. The x-rays would illuminate the 2.2 mm diameter target capsule, ablating its surface and compressing the frozen DT fuel to ultrahigh density. This “indirect drive” may

provide better compression than the “direct drive,” where the laser beams interact directly with the target surface, but the Hohlraum targets are more complex and more difficult to inject into a chamber and align with the laser beams.

Table 8 shows that many issues must be resolved before ICF reactors can become practical.

Potential solutions are available for most issues, but technology development has not yet been completed, especially for efficient, reliable, cost-effective drivers.



Nuclear Fusion. Figure 24

The National Ignition Facility, USA. 192 laser beams (in red tubes) go to the target chamber (Courtesy of Lawrence Livermore National Laboratory)



Nuclear Fusion. Figure 25

The NIF target chamber, with holes where the laser beams enter. (The wide angle photo makes the 10-m diameter chamber appear to be larger.) (Courtesy of Lawrence Livermore National Laboratory)

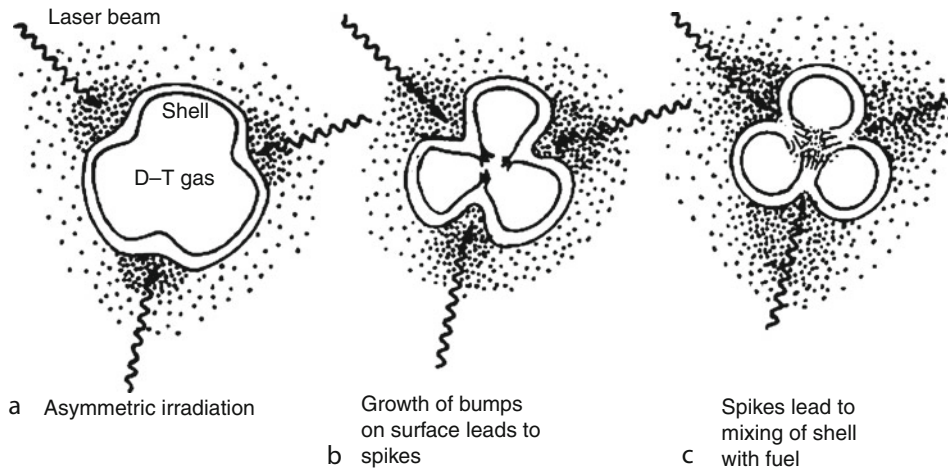
Table 9 lists some issues of target design and manufacture.

Laser Development

Great improvements are needed in laser efficiency, pulse repetition rate, and target injection rates. In the USA, two experiments are aimed at high average power lasers: Electra and Mercury.

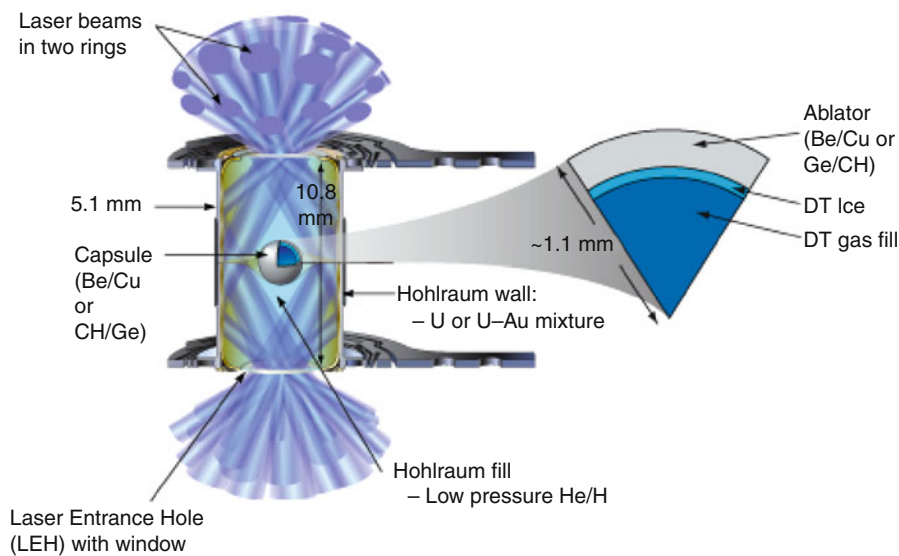
The *Electra* KrF laser experiment at the Naval Research Laboratory is illustrated in Fig. 28.

Electron beams in vacuum from the cathode pass through the foil into the Kr and F₂ gases in the laser cell, where they cause excitation, resulting in a laser beam that exits through the amplifier window. *Electra* will run at 5 Hz with a laser output of 400–700 J. This will be large enough to develop technologies that can be scalable to the 50–150 kJ needed for a fusion power plant beam line.



Nuclear Fusion. Figure 26

Growth of the Rayleigh-Taylor instability, causing breakup of the target shell (exaggerated for clarity)



Nuclear Fusion. Figure 27

A Hohlraum target (Courtesy of Lawrence Livermore National Laboratory)

Figure 29 shows the *Mercurydiode-pumped solid state laser*.

The material in which light amplification occurs is crystalline strontium fluoro-apatite doped with ytterbium (Yb:S-FAP), which has a long excited state lifetime and high saturation fluence. High-power diodes inject 900 nm wavelength light into the Yb:S-FAP, which emits near 1,047 nm. About 15–20% of the

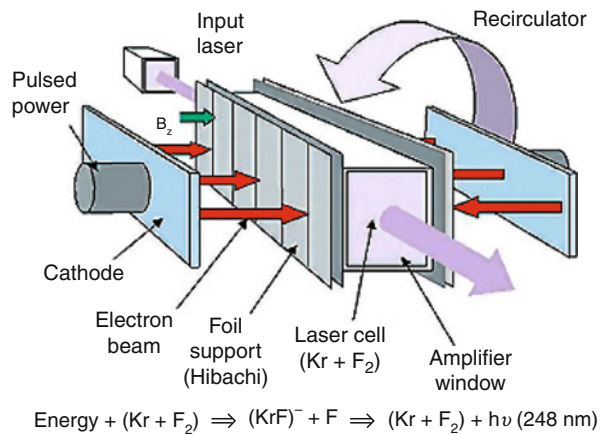
diode energy is dissipated as heat in the Yb:S-FAP, which is removed by turbulently flowing helium gas [14]. The Mercury laser should deliver 100 J pulses at a rate of 10 Hz. If successful, this could be scaled up by using many units to MJ energies for a power plant. With mass production of diodes, their cost may fall to pennies per Watt, as needed for an economical power plant.

Nuclear Fusion. Table 8 Technical issues of ICF reactor development

Theory	Understand the physics of energy absorption, reflection, heat transport, compression, instabilities ignition, and burn
Experiments	Attain satisfactory values of the critical elements for high energy gain <ul style="list-style-type: none"> • Beam-to-fuel coupling efficiency • Avoiding fuel preheating before compression is complete • Implosion symmetry • Driver energy per unit fuel mass
Target manufacture	Develop an automated system to manufacture high-gain spherical targets accurately and cheaply. Select target materials to avoid production of long-lived radioisotopes. <p>Use cryogenic refrigeration to keep the DT fuel frozen.</p>
Target shelf life	Tritium decay heat can damage the cryogenic target, so targets have a finite shelf life.
Target injection	Develop a target injection and guidance system to ensure that the target is at the focus of the beams when the driver is fired.
Diagnostics	Measure the parameters of experiments, such as implosion velocity, density distributions, energy distributions, ablated matter, laser beam reflection, implosion symmetry, and fusion reaction products, with fine spatial and temporal resolution.
Chambers	Develop reactor chambers to withstand repeated ($>3 \times 10^8$) explosions without failure. (A 1,000 MJ yield has the explosive energy of 240 kg of TNT, but less momentum).
Chamber Clearing	Remove debris and gases rapidly between explosions, in order to avoid attenuating the incident driver beams.
Drivers	Develop laser or ion beams with high-energy (1–10 MJ), high-pulse repetition rate (1–10 Hz), proper pulse shape (duration 10–20 ns), suitable wavelength (lasers), good efficiency ($\geq 8\%$), and low enough cost.
Reliability	Develop power supplies, power conditioning equipment, diodes, and optical components to operate reliably for $>10^9$ shots.

Nuclear Fusion. Table 9 Target design and manufacture issues

Hydrodynamic stability and high compression
Good sphericity
Fast fill and layering
Uniform thickness shells
Very smooth surfaces
Minimum preheat of fuel
High-Z material to stop electrons and x-rays
Minimum production of long-lived radioisotopes
Easy handling and storage
Capable of high-velocity acceleration and injection without damage
Cheap materials, like plastics
Simple, inexpensive fabrication
Target cost goal typically <10 cents per target

**Nuclear Fusion. Figure 28**

The Electra KrF laser (Courtesy of Naval Research Laboratory)

Hypervelocity Impact Fusion Small pellets (mass ~ 1 g) would be accelerated to velocities $>10^5$ m/s and shot at a solid target. Either the projectile or the target would contain DT fuel. During impact, intense compression and shock heating would occur, igniting some of the fuel. The compression could be

made nearly three-dimensional by shaping the projectile and target, Fig. 30.

If compression is great enough, a high energy gain ratio $Q > 30$ may be attained. The fusion yield should be low enough that the blast chamber is not destroyed and high enough for economical power generation. The required projectile energy is about 25 MJ. Possible means of acceleration include electromagnetic rail guns, electrostatic accelerator, light gas guns, traveling electromagnetic wave accelerator, and laser ablation.

Challenges of Plasma Theory

In order to model fusion reactor behavior plasma physics theory must take into account many phenomena, Table 10.

Many of these models are based on simplifications of the Boltzmann transport theory and Maxwell's equations. Because these phenomena are diverse and often involve nonlinear differential equations, there is no unified theory that encompasses all of them. Instead,

these phenomena may be modeled separately and the models stitched together to get an approximate picture of plasma behavior. Analytic theories are useful to understand how individual phenomena vary with input parameters, but are limited by the simplifying assumptions that are needed to make the equations solvable.

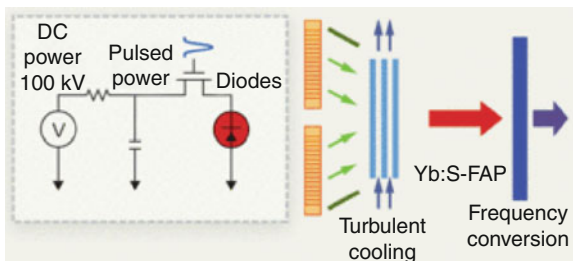
Large-scale computer simulations are widely used to model these effects. Three-dimensional models can be used to predict plasma behavior. The successful modeling of many phenomena lends confidence to predictions of the behavior of future experiments, such as ITER, but experimental data will still be needed to confirm the predictions.

Fusion Technology Issues

Many areas of technology are required for successful development of fusion power, Table 11.

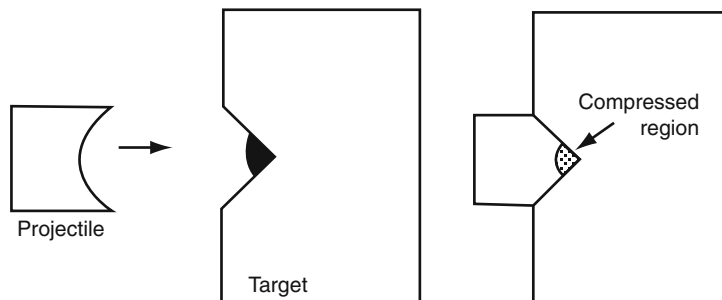
Most of these issues have been successfully resolved during the 6 decades of fusion research. The ITER experiment should demonstrate the successful combination of many of these technologies, but will not demonstrate much tritium breeding, durable wall materials, high availability, or electrical power generation.

ITER must deal with transient heat loads caused by "edge localized modes" (ELMs) and disruptions, which can potentially damage the first wall or divertor. A robotic system to replace damaged wall tiles will be tested under high radiation levels in ITER. Development of first wall or divertor materials that can resist high heat fluxes and are resistant to damage by 14-MeV neutrons will not be completed by ITER, which uses more conventional structural materials.



Nuclear Fusion. Figure 29

The Mercury diode-pumped solid state laser (Courtesy of the Lawrence Livermore National Laboratory)



Nuclear Fusion. Figure 30

Hypervelocity impact projectile and target shaped to provide good compression

Nuclear Fusion. Table 10 Some phenomena to be modeled by plasma theory

Magnetohydrodynamic (MHD) instabilities	Motion of plasma boundary, driven by plasma pressure and Lorentz force
Radiation losses	Line radiation, bremsstrahlung, cyclotron radiation
Coulomb collision effects on velocity distributions	Boltzmann or Fokker-Planck equation
Plasma wave generation, cutoffs, and resonances	Maxwell equations coupled to plasma current density and magnetic field
Wave-particle interactions and microinstabilities	Anisotropic or non-Maxwellian distributions; kinetic theory (based on Boltzmann equation)
Nuclear fusion reactions	Cross sections averaged over ion velocity distributions
Particle and energy transport	Transport coefficients in conservation equations; scaling laws derived from experimental data
Plasma-wall interactions	Sputtering, vaporization
Impurity transport and control	Edge plasma and divertor theory
Plasma heating and current drive	Neutral beam injection, wave heating

In parallel with the ITER project, the International Fusion Materials Irradiation Facility (IFMIF), planned to be built in Japan, will bombard a flowing lithium target with 40 MeV deuteron beams to produce an intense flux of energetic neutrons for testing small samples of candidate fusion reactor materials. [Figure 31](#) shows a drawing of IFMIF [15].

Use of high-temperature superconductors is becoming feasible for fusion devices [16]. They could operate near liquid nitrogen temperature (77 K), instead of near liquid helium temperature (4 K). This would simplify the design of the coils, cryostats, and instrumentation, and greatly reduce the refrigeration power and cost of refrigeration systems.

Coating the plasma chamber walls with lithium may help improve plasma performance. Usually,

many of the deuterons escaping from the plasma hit the wall, where they are neutralized. Some of them re-emerge from the wall as cold neutral gas, which keeps the plasma edge temperature low, making a steep temperature gradient, which causes rapid heat conduction. Lithium absorbs escaping deuterium, reducing back-flow of cold deuterium gas, and flattening the temperature profile, which can reduce plasma electron turbulence and heat conduction rates [17].

Fusion Reactor Design Studies

Electric power utility companies have many criteria for power plant selection and siting [18] including:

- Output of power plant, compatibility with power grid
- Grid stability
- Required transmission lines
- Capital cost per kW
- Simplicity of plant
- Construction time
- Reliability and availability of power plant
- Ease and speed of maintenance
- Plant lifetime
- Load-following (ability to vary output to meet demand)
- Seismic activity and flood danger
- Availability of fuel, and fuel cycle costs
- Decommissioning costs
- Adequacy of resources, such as helium, lithium, and niobium
- Feasibility of shipping large, heavy components
- Staff numbers, skills, and costs
- Security requirements and costs
- Government laws and regulations
- Ease of licensing
- Intervener lawsuits
- Safety
- Environmental issues
- Plant emissions and waste products
- Willingness of financial institutions to invest
- Public acceptance

The availability of fusion power plants is difficult to predict, because current experiments often have equipment failure rates that would be unacceptable to utilities. The availability of some systems can sometimes be

Nuclear Fusion. Table 11 Fusion technology issues

Component	Issues
Magnet coil systems	High magnetic field, superconducting cable, stresses, quench protection
Plasma heating, fueling, current drive	Plasma compression Magnetic induction Electromagnetic waves Particle beam injection Plasma guns
First wall	Low-Z (atomic number), high heat flux, heat removal, radiation damage, tritium trapping
Blanket, shield	Tritium breeding ratio, neutron and gamma shielding, heat removal, radiation damage
Divertors (channels that remove plasma impurities)	High heat flux, sputtering, vaporization, tritium trapping, vacuum pumping
Energy conversion	Efficient conversion of fusion energy into electrical energy, such as with a steam cycle or helium gas turbine. High efficiency requires high coolant temperature, which causes materials problems
Materials issues	Stress, embrittlement, swelling, creep, fatigue, tritium trapping
Vacuum systems	Pumping speed, conductance, cryopanel
Cryogenic systems	Coolant flow, refrigeration, liquefaction, heat shields
Plasma diagnostic systems	Space-time mapping of magnetic field, electric field, plasma density, temperature, impurities, waves
Control systems	Plasma control, heating, current drive, magnets, fusion burn, electrical power plant
Maintenance	Magnets, first wall, heating systems, diagnostics, heat transfer systems
Safety and environmental systems	Routine emissions, accident scenarios, decommissioning, waste disposal
Power plant designs	Criteria for attractive power plants Reliability, availability, and maintenance Economics Fusion–fission hybrids

improved by providing redundant components, with slight increases in cost. Repair of large superconducting magnets would be difficult, if not impossible, so designers try to ensure that they will last for the life of the plant.

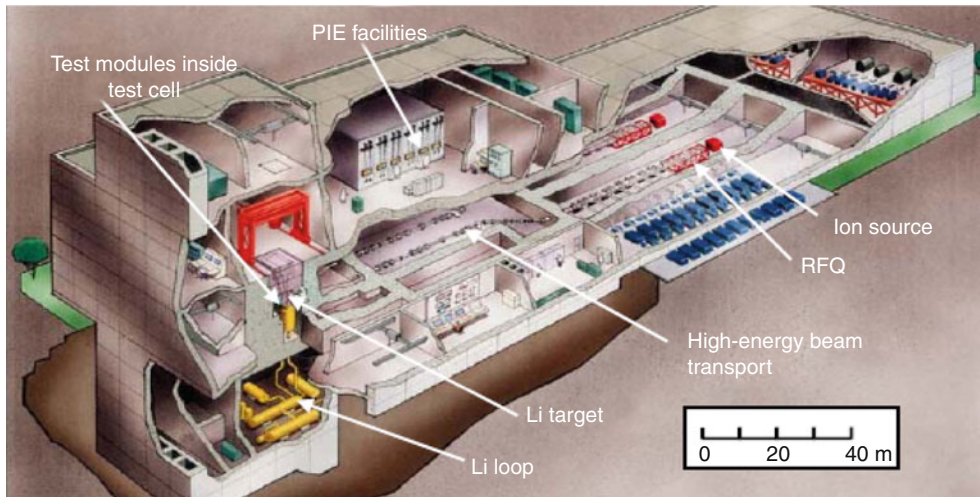
Fusion reactors have an economy of scale, which means that a 3,000 MWe power plant would probably have a much lower COE than a 300 MWe plant, but utilities may not wish to build very large power plants, to avoid large grid perturbations during outages [19].

The electricity from a fusion power plant could also be used to make hydrogen by high-temperature electrolysis. The hydrogen could be used as fuel for

transportation and industrial processes. If the fusion reactor blanket could operate at $T > 800^{\circ}\text{C}$, then the heat could be used directly for hydrogen production by a thermo-chemical system, such as one using a sulfur-iodine cycle [20].

Fusion–Fission Hybrids

Present light water fission reactors burn mainly the fissile isotope ^{235}U , which is only 0.711% of natural uranium. (“Fissile” means that it can be fissioned easily by slow neutrons.) The other 99.27% is mostly ^{238}U , which does not fission easily. A strong neutron source is needed in order to convert the ^{238}U into ^{239}Pu , which is



Nuclear Fusion. Figure 31

The international fusion materials irradiation test facility (Courtesy of the International Energy Agency)

another fissile fuel, if we are to generate energy from the other 99% of the uranium resources.

That strong neutron source could be a “fast” fission reactor, such as a liquid metal fast breeder reactor, but it takes years for such a reactor to generate enough fuel to start up a second reactor. The neutron source could be a high-power accelerator beam (like IFMIF), but fusion plasma neutron sources would be better where large quantities of materials are to be irradiated.

A fusion–fission hybrid would use uranium or thorium in the fusion reactor blanket, in addition to lithium. The blanket could be optimized either to produce more heat and electricity from the hybrid plant or to breed fissile fuel (^{239}Pu or ^{233}U) for use in satellite fission reactors. Hybrid blanket design studies indicate that one hybrid reactor could provide fuel for about five light water reactors. An “Energy Park” could have a fusion reactor plus several fission reactors and a fission reactor fuel reprocessing plant. With modern reactor designs, much of the spent fuel radioactive wastes could be recycled in the fission reactors or “incinerated” (transmuted to stable isotopes) in the fusion reactor blanket [21].

Future Directions

Most past magnetic fusion experiments have had small minor radii, so energy confinement times were short

(<0.1 s), and impurities from the wall prevented optimum performance. The fusion research field needs bold investment to build large, high-power devices, such as ITER, in several other areas:

- Large compact toroid experiments (FRC and spheromak)
- A high-power accelerator neutron source for materials testing
- A plasma-based volumetric neutron source for materials testing
- Tokamak demonstration power plant construction during ITER operation
- Two reactor-sized stellarators following LHD and W 7-X
- A reactor-sized tandem mirror following Gamma-10
- Rapid-pulse, high-gain laser fusion demonstration power plants
- Full-scale heavy ion beam driver demonstration

Fusion reactor design teams in Europe, Japan, and the USA have done many design studies of tokamaks, stellarators, inertial fusion, and other types of fusion power plants. The capital costs of most fusion reactors are very high, but the fuel costs are very low. The resulting cost of electricity (COE) is typically predicted to be about 1.5–2 times as high as the COE from coal plants and nuclear fission reactors. Fusion power plants could become competitive

- If cheaper fusion reactors were developed, such as from compact toroids
- If fusion–fission hybrids were developed
- If carbon emissions were taxed
- If public opposition made it difficult to build new fission reactors or coal power plants

Acknowledgments

The following provided helpful comments on this article: Ralph Moir, Chan Choi, Lee Cadwallader, Nicholas Tsoulfanidis, Alex Parrish, and the Institute for Plasma Research (Gandhinagar, India). Charlou Dolan drew many of the figures.

Bibliography

Primary Literature

1. Jaskula BW (2010) Lithium. USGS 2008 Minerals Yearbook, United States Geological Survey
2. Freidberg J (2006) Plasma physics and fusion energy. Cambridge University Press, Cambridge
3. "All-the-world's tokamaks" (2010) <http://www.toodlepip.com/tokamak>
4. Bolt H (2001) Materials for Fusion. European White Book on Fundamental Research in Materials Science, Max-Planck-Gesellschaft, Munich, Section 2.9, Figure 2.17. (See also <http://www.mpg.de/bilderBerichteDokumente/dokumentation/europWhiteBook/>)
5. Hazeltine R et al (2009) Research needs for magnetic fusion energy sciences. Report of the Research Needs Workshop (ReNeW), Bethesda, 8–12 June, 2009, U.S. Department of Energy, p 203
6. Hazeltine R et al (2009) Research needs for magnetic fusion energy sciences. Report of the Research Needs Workshop (ReNeW), Bethesda, 8–12 June, 2009, U.S. Department of Energy, p 213
7. Tacetti JM et al (2003) FRX-L: a field-reversed configuration plasma injector for magnetized target fusion. *Rev Sci Instrum* 74(10):4314–4323
8. Slough J et al (2007) The pulsed high density experiment: concept, design, and initial results. *J Fusion Energy* 26:199–205
9. Ichimura M et al (2006) CRF experiments and potential formation on the GAMMA 10 tandem mirror. *Plasma Sci Technol* 8(1):87–90
10. Dolan TJ (1994) Magnetic electrostatic plasma confinement. *Plasma Phys Controlled Fusion* 36:1539–1593
11. Boxer AC et al (2010) Turbulent inward pinch of plasma confined by a levitated dipole magnet. *Nature-Physics* 6:207–212, Supplementary Information
12. Storms E (2007) The science of low energy nuclear reaction: a comprehensive compilation of evidence and explanations. World Scientific, Hackensack
13. Dolan TJ (1982) Fusion research. Pergamon, Elmsford, Chapter 15
14. Ebberts C, Caird J, Moses E (2009) *Laser Focus World* 45, No. 3, March 1
15. IFMIF International Team (2004) International Fusion Materials Irradiation Facility (IFMIF) Comprehensive Design Report. International Energy Agency, Paris
16. Hazeltine R et al (2009) Research needs for magnetic fusion energy sciences. Report of the Research Needs Workshop (ReNeW), Bethesda, 8–12 June, 2009, U.S. Department of Energy, Thrust 7, pp 285–293
17. Zakharkov LE et al (2004) Ignited spherical tokamaks. *Fusion Eng Des* 72:149–168
18. Kaslow J et al (1994) Criteria for practical fusion power systems: report from the EPRI fusion panel. *J Fusion Energy* 13(2/3):181–183
19. Dolan TJ (1993) Fusion power economy of scale. *Fusion Technol* 24:97–111
20. Terada A et al (2007) Development of hydrogen production technology by thermochemical water splitting IS process, pilot test plan. *J Nucl Sci Technol* 44(3):477–482
21. Freidberg J et al (2009) Research needs for fusion-fission hybrid systems. Report of the Research Needs Workshop (ReNeW), Gaithersburg, Maryland, September 30–October 2, 2009, U.S. Department of Energy

Books and Reviews

- Braams CM, Stott PE (2002) Nuclear fusion: Half a century of magnetic confinement fusion research. Institute of Physics, Philadelphia
- Chen FF (1984) Introduction to plasma physics and controlled thermonuclear fusion. Plenum, New York
- Dinklage A et al (2005) Plasma physics – confinement, transport, and collective effects. Springer, Berlin
- Dolan TJ (1982) Fusion research. Pergamon, Elmsford
- Freidberg J (2006) Plasma physics and fusion energy. Cambridge University Press, Cambridge

Nuclear Power, Economics of

M. R. DEINERT

Department of Mechanical Engineering,
The University of Texas at Austin, Austin, TX, USA

Article Outline

Glossary
Definition of the Subject
Introduction

Present and Future Value of Money
 Levelized Costs
 Lead and Lag Times in Cost Calculations
 Estimating Uncertainties
 Costs and Their Uncertainties
 Discount and Interest Rates
 Cost Comparisons and External Costs
 Future Directions
 Acknowledgments
 Bibliography

Glossary

c Onetime costs [\$]
 E_c Cost of electricity [\$]
FV Future value [\$]
HM Heavy metal, refers to the uranium and or trans-uranic component of fuel
kg Kilogram
kW Kilowatt
kWh Kilowatt hour
kWh(e) Kilowatt hour electric
mill \$0.001
n Number of years
MOX Mixed oxide fuel
MWh(e) Megawatt hour electric
p(t) Distributed costs [\$/year]
Pu Plutonium
PV Present value [\$]
r Yearly rate of return in discrete discounting
 $r(x_i, x_j)$ Correlation coefficient
SD Standard deviation
SF Spent fuel
SW Separative work
t Time [year]
U Uranium
UOX Uranium dioxide fuel
VHLW Vitrified high-level waste
 x_i Denotes a cost component
 α Linear cost escalation rate [\$/year²]
 β Linear cost escalation intercept [\$/year]
 δ Delta function
 ρ Discount or interest rate [1/year]

Definition of the Subject

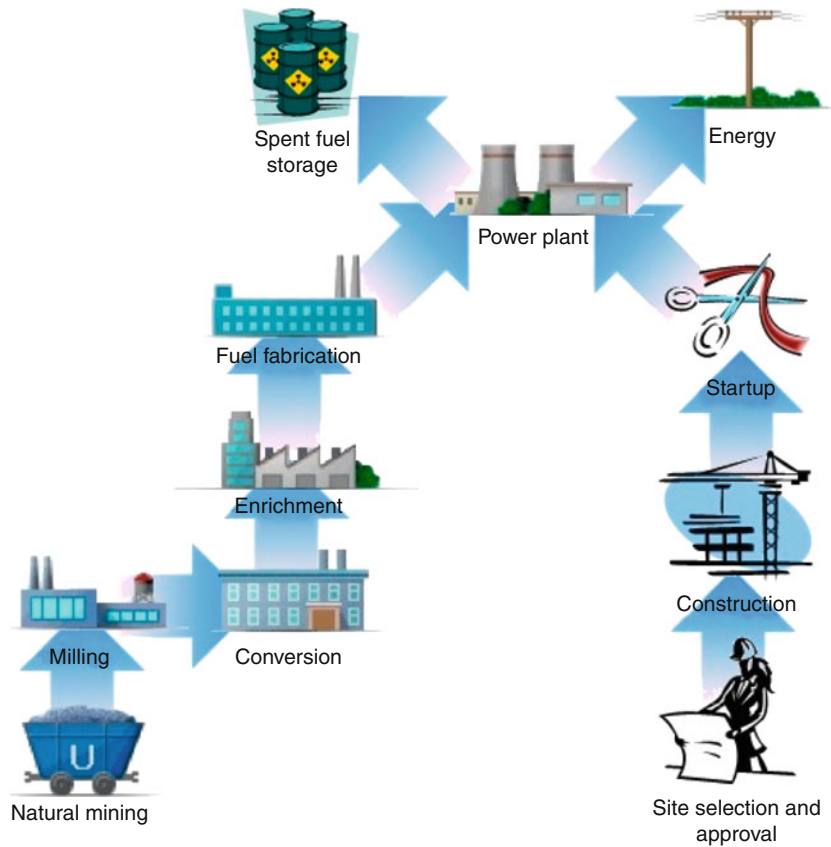
Financial viability is an important consideration when deciding whether to proceed with any large-scale

engineering project. Many studies of nuclear power economics have been undertaken in an attempt to predict its overall costs or competitiveness (e.g., [1–4]). While these studies tend to differ in their assumptions about construction and operating expenses, they all use similar frame works for the analysis. In essence, the idea is to predict the total cost of producing electric power over the lifetime of a facility and compare that to the market value of the electricity produced. All other things being equal, the larger the ratio of revenue to cost the better the project.

Introduction

Economic assessments of nuclear power tend to be complicated, and not just because of the number of components that have to be factored in, Fig. 1. The costs of any large project also depend on how it is financed, and whether this is done through the issuing of bonds by the entity undertaking the project, borrowed money, allocation of liquid assets, and/or the use of complex financial instruments such as derivatives. Tax rates, both federal and local, can also play an important role. The revenue stream from a nuclear power facility itself depends on whether the markets into which the electricity is fed are regulated or unregulated and whether arrangements exist that involve the sale of electricity at fixed rates to municipalities where the facilities reside. The effect of hard to predict market forces can affect the price of electricity itself.

In addition, nuclear power has some peculiarities that are unique to the industry. Among these is the need to safely store the radioactive waste products that are contained in spent nuclear fuel, for extended periods of time. At present, only the US Waste Isolation Pilot Plant, in Carlsbad NM, is actively used for this purpose, and then only for transuranics (a very specific class of high-level waste) from the US nuclear weapons program. Because of this, *interim* storage facilities exist at civilian nuclear power facilities to handle spent nuclear fuel, but the final disposal costs remain uncertain (e.g., [2]). In addition, most modern reactors run on a uranium fuel that has been enriched to contain more of the uranium 235 isotope than is present in natural uranium (natural uranium is made up of two isotopes: ²³⁸U and ²³⁵U. The ²³⁸U isotopes comprises



Nuclear Power, Economics of. Figure 1

Components of a nuclear power facility. Construction of a nuclear power facility begins with site selection and approval. Facility construction (reactor and related facilities) comes next, with final startup typically coming 6–12 years after the start of construction activities. The process of producing uranium fuel typically starts 2 years prior to fuel placement in a reactor with uranium mining, milling, conversion enrichment, and fuel fabrication. Finally, fuel is transported to the reactor site. The fuel pipeline will function continuously as the reactor operates. Interim on-site storage is common until permanent storage facilities are established (Figure courtesy of Robert Bell, The University of Texas at Austin)

99.3% of the uranium atoms and ^{235}U 0.7%. It is the ^{235}U isotope that readily undergoes fission in the thermal reactors that dominate the civilian nuclear power industry.). Because this isotope can also be used in the manufacture of nuclear weapons, concerted efforts have been undertaken to limit the development of enrichment facilities outside of a few highly industrialized countries (e.g., [5, 6]). This of course has the effect of forcing nuclear power facilities in less industrialized countries to buy fuel from external sources. Consequently, a common concern is that the fuel supply could be restricted for political reasons (e.g., [7]).

Uncertainties in fuel availability and spent fuel disposal introduce an element of financial risk that is difficult to quantify. Another potential complication comes from uncertainty in whether construction of a nuclear power plant will actually lead to an operational facility. It has in fact happened that facilities have been undertaken whose construction has been halted, or which have been completed but closed before becoming commercially operational. The Shoreham Nuclear Power Plant in Long Island, NY, is an example of this. The facility was completed in the mid-1980s, but public protest resulted in its closure before it produced commercial power.

Nuclear Power, Economics of. Table 1 *Life-cycle cost estimates for electricity produced from nuclear, coal, and gas.* The different costs of electricity reflect different life-cycle assumptions between the two studies [4, 11]. Costs were adjusted to \$2,007

Production	MIT [4] [cents/kWh(e)]	University of Chicago [11] [cents/kWh(e)]
Nuclear	8.4	5.6
Coal	6.2	4.5
Gas	6.5	3.4

Because of the above factors, and others, it is difficult to make blanket statements about the economics of nuclear power, and facilities need to be evaluated on an individual basis. However, several recent studies have taken a look at nuclear power relative to competing technologies, see Table 1, with somewhat mixed results that reflect the different assumptions that went into the analyses. In the current article, the focus is largely on a discussion of how the cost of a nuclear power facility would be computed when taking into consideration factors such as its construction, operation, maintenance, and decommissioning.

Many of the elements that contribute to the life-cycle cost of a nuclear power plant are common to other types of facilities as well. These include:

Capital costs – The costs associated with building the plant and its components.

O&M – The cost to operate and maintain facilities.

Depreciation – This is a charge recorded against earnings that takes into consideration the lifetime of capital components and the fact that they must be replaced as a part of operating expenses. Depreciation of capital components is typically added to the yearly operations and maintenance costs.

Interest – The money paid for the use of borrowed capital or for bonds that have been issued.

Taxes – Both federal and local may apply.

Interest rate – The annual amount of money paid to a lender or bond holder for the use of capital as a percentage of the amount to be repaid.

Discount rate – Effectively the same as the interest rate, see section on discount rates.

Several other factors that affect the cost of nuclear power are particular to the industry. These include:

Fuel costs – A peculiarity of nuclear power facilities is that they do not always buy the fuel that is used to run them. Instead, they will often lease it, sometimes from a company that was created explicitly for the purpose of doing this. The reason has to do with accounting practices, and it is sometimes cost effective for a utility to acquire its fuel in this way [8].

Nuclear waste fund fee – This is a fee levied in the United States to cover the Federal Government's obligation to take possession of spent nuclear fuel and dispose of it [8]. The current rate is \$0.001/kWh of electricity produced, which is paid into the Nuclear Waste Fund (The Nuclear Waste Fund was established under US Code Title 42, Chapter 108, Subchapter III, 10222 – i.e., The Nuclear Waste Policy Act of 1983). This monetary unit is so common within the utility industry that it is often given a special unit called a “mill,” where 1 mill = \$0.001 [8].

On site spent fuel storage cost – The cost to store spent nuclear fuel on site.

Decommissioning costs – The costs associated with removing the power plant and its components along with returning the site to an unrestricted use.

By far the dominant cost for a typical nuclear power plant is that of constructing the facility itself [4]. From the perspective of the people who are building a facility, the cash flows that are required for construction depend not only on the facility cost, and the time required to build it, but also on how it is financed. For example, cash flows associated with a facility financed with cash will be different than those for one financed with a loan that is to be paid back over a fixed period of time, or a bond. Once a facility is built, there are operation and maintenance charges to keep it in working order, and these are spread out over the operating life of the facility. Taxes on property begin as soon as land is acquired, and those on the facility itself depend on the location and municipality, but are likely to be yearly as are taxes on income. Depreciation too is spread out over the operating life of a component (a discussion of common depreciation methods can be found in [4]), and decommissioning costs are

incurred at the end of a facility's life. However, it is typical that money must be set aside for decommission costs well in advance of them.

Present and Future Value of Money

The variable nature of the cash inflows and outflows complicates the cost analysis for any industrial facility, but especially those with long lifetimes as is typical with nuclear power plants. When undertaking a life-cycle cost study, it is also important to recognize that not all money is equivalent. Specifically, a dollar paid or earned today has a different value than does one paid or earned a year from now. The reason for this stems from the fact that money can devalue due to inflation, but also because money held today can be invested and earn interest. Conversely, money received in the future has less value, both because it may have devalued but also because of the inability to invest it until it is received. These concepts are captured in what is often referred to as the "time value of money" [9]. For example, \$100 invested today, at a yearly return of 5%, will yield \$105 in 1 year. In the parlance of financial engineering, \$100 invested this way has a *future value* of \$105, and \$105 received in 1 year has a *present value* of \$100. From the perspective of an investor who expects a yearly return of 5%, both are equivalent.

The relationship between *future value* (FV) and *present value* (PV) is simple:

$$FV = PV(1 + r)^n \quad (1)$$

Here r is the yearly rate of return, and n is the number of years over which the investment takes place (not necessarily an integer), and PV is said to be *compounded* over n years. Conversely, the *present value* of FV (received in n years) is given by:

$$PV = FV/(1 + r)^n \quad (2)$$

An important point that is often overlooked is that $PV = FV$ when the expected rate of return is zero. This situation can effectively arise in environments, where the rate of return is equal to the inflation rate. Equation 2 is an example of *discounting*. When the effects of inflation are taken into consideration Eqs 1, 2 become:

$$FV = PV(1 + r - i)^n \quad (3)$$

and

$$PV = FV/(1 + r - i)^n \quad (4)$$

Here i is the yearly inflation rate. Equations 3, 4 give the present and future values in constant, or inflation adjusted, dollars. The quantity $\rho = r - i$ is referred to as the "real rate of return" as opposed to the "nominal rate of return" r , and ρ , r , and i all range between 0 and 1. Most economic analyses are done assuming a real rate of return, and that convention is adopted here.

Equations 1–4 are discrete representations of simple compounding and discounting, and there are several very good descriptions of how they can be extended to more complex situations in general (e.g., [10]), and to the nuclear power industry in particular [1, 8]. However, as these references quickly show, applying discrete financial models becomes cumbersome when the system being analyzed is complex. An alternative is to use continuous approximations to Eqs 3, 4, which then become:

$$FV = PVe^{\rho t} \quad (5)$$

and

$$PV = FVe^{-\rho t} \quad (6)$$

Here ρ is again the rate of real rate of return per unit time t .

Levelized Costs

Nuclear power facilities typically operate over extended time periods, and some in the United States are even in the process of receiving license extensions that will bring their operating life spans to 60 years (e.g., [11]). As a result, the present value of funds that are used to build, maintain, fuel, or decommission a nuclear facility will depend significantly on when the costs occur. Therefore, the only way to accurately estimate the total cost of a facility is to compute its total *present value* relative to a specific date (usually the date of startup of the facility). This process is called *levelizing*. In fact, in the United States, all large-scale government projects are required to perform this type of levelized cost analysis [12]. Expenses that take place before the reference date have what is called a *lead time* as they happen *ahead of* the reference time. Expenses that take place after the reference time have a *lag time* as they happen behind the reference time [1, 13]. A framework for how

to levelize the costs for nuclear power facilities in terms of lead and lag times was given in a study done by the Organization for Economic Cooperation and Development (OECD) in collaboration with the Nuclear Energy Agency (NEA) [1]. Continuous discounting can be used within this framework to levelize all costs, as well as the revenue from electricity production, to the date at which fresh fuel is loaded into a reactor. Continuous discounting is mathematically simpler than its discrete alternative and often introduces negligible error relative to the large variances for unit cost predictions.

With this approach, levelized costs are easily obtained by multiplying the time-dependent costs, $p(t)$, by the discounting factor, $e^{-\rho(t)t}$, and integrating the product:

$$PV = \int_{t_1}^{t_2} p(t) e^{-\rho(t)t} dt \quad (7)$$

where PV [\$] is again the present value of a cash outflow, $p(t)$ [\$/year], ρt is the real interest or discount rate [1/year], and t is time [year]. It is common in many studies to assume that ρt is constant, though this is not necessary. The total cost of a system over its life is given by the sum over the costs incurred from its construction to its decommissioning:

$$E_c = \sum_i PV_i \quad (8)$$

The life-cycle cost in \$/kWh(e) is calculated by dividing E_c by the total kWh's of electricity produced. Some economists argue that the electricity production should itself be discounted using Eq. 7 to factor in the temporal nature of the revenue stream (e.g., [14]). It might seem counter intuitive that a unit of energy, being immutable, should be discounted. However, economists who advocate this approach argue that energy (like money) can be viewed as having greater value when it is available in the near term than if its availability is off in the distance. Another interpretation was given by Hannon [14] who suggested that the energy discount rate reflects a society's desire to convert "a present surplus energy into an energy-transformation process so that a greater surplus of energy can be created in the future, rather than consuming the energy now." While energy discounting is used (e.g., [4, 8]), the approach is not universal.

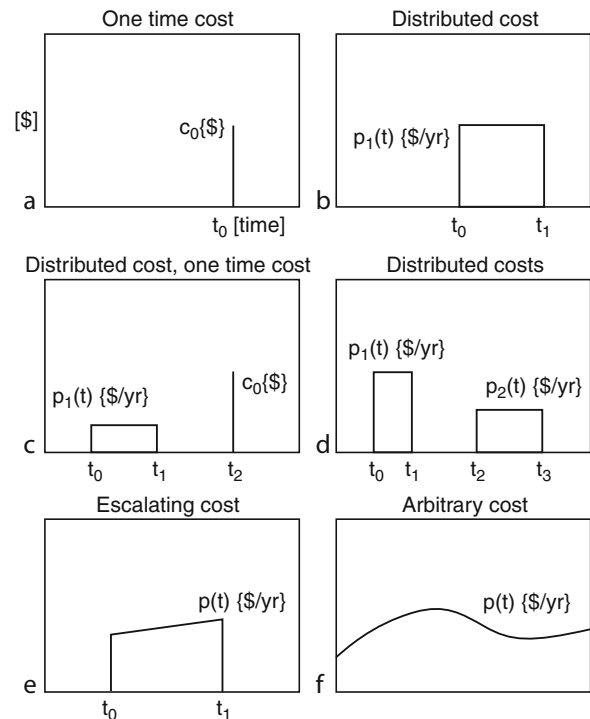
Examples. Figure 2 shows several distributions for $p(t)$ that are relevant to nuclear power systems and out of which any cost or revenue distribution can be built. Equation 7 is easy to apply to each situation when it is done systematically.

- (a) The present value of a onetime cost, c_0 , that occurs at time, t_0 , is easy to derive using Eq. 7. One starts by noting that:

$$\begin{aligned} p &= 0 & 0 \leq t < t_0 \\ p &= c_0 & t = t_0 \\ p &= 0 & t > t_0 \end{aligned} \quad (9)$$

Assuming a constant discount rate, Eq. 7 becomes:

$$PV = \int_0^{t_0} c_0 \delta(t_0) e^{-\rho t} dt \quad (10)$$



Nuclear Power, Economics of. Figure 2

Common payment, cost, and revenue distributions. Common distributions for $p(t)$ applicable to Eq. 7 are shown. In the figure, c_0 stands for one time cost, and $p(t)$ for distributed cost. The left-hand side of the time axis is assumed to represent $t = 0$

where $\delta(t_0)$ is the delta function (zero everywhere but t_0). Equation 10 then just gives:

$$PV = ce^{-\rho t_0} \quad (11)$$

which is just Eq. 6 again.

- (b) The present value of a uniformly distributed cost is only slightly more complicated to calculate. Here:

$$\begin{aligned} p &= 0 & 0 \leq t < t_0 \\ p &= c_1 & t_0 \leq t \leq t_1 \\ p &= 0 & t > t_1 \end{aligned} \quad (12)$$

Equation 7 then becomes:

$$PV = \int_{t_0}^{t_1} c_1 e^{-\rho t} dt \quad (13)$$

which has the simple solution:

$$PV = \frac{c_1 [e^{-\rho t_0} - e^{-\rho t_1}]}{\rho} \quad (14)$$

- (c) The present value of a distributed cost and a onetime cost is just the sum of the results given in Eqs. 11, 14:

$$PV = c_0 e^{-\rho t_0} + \frac{c_1 [e^{-\rho t_0} - e^{-\rho t_1}]}{\rho} \quad (15)$$

It is implicit in Eq. 15 that the discount rate is the same for both distributions; however, this is not always the case, and not a necessary restriction.

- (d) The present value of two uniformly distributed costs is just obtained using the result in Eq. 14 twice:

$$PV = \frac{c_1 [e^{-\rho t_0} - e^{-\rho t_1}]}{\rho} + \frac{c_2 [e^{-\rho t_2} - e^{-\rho t_3}]}{\rho} \quad (16)$$

Here c_1 and c_2 are the uniform cost rates for the respective distributions. It is again implicit that the discount rate is the same for both distributions, though this is again not a necessary restriction.

- (e) Linearly escalating costs are also easy to deal with. Here the cost function is given by:

$$\begin{aligned} p &= 0 & 0 \leq t < t_0 \\ p &= \alpha t + \beta & t_0 \leq t \leq t_1 \\ p &= 0 & t > t_1 \end{aligned} \quad (17)$$

where α has units of [\$/year²] and β [\$/year]. In this case, Eq. 7 becomes:

$$PV = \int_{t_0}^{t_1} (\alpha t + \beta) e^{-\rho t} dt \quad (18)$$

Equation 18 has the simple solution:

$$PV = \alpha \left(\frac{t}{\rho} - \frac{1}{\rho^2} \right) (e^{-\rho t_1} - e^{-\rho t_0}) + \frac{\beta}{\rho} (e^{-\rho t_1} - e^{-\rho t_0}) \quad (19)$$

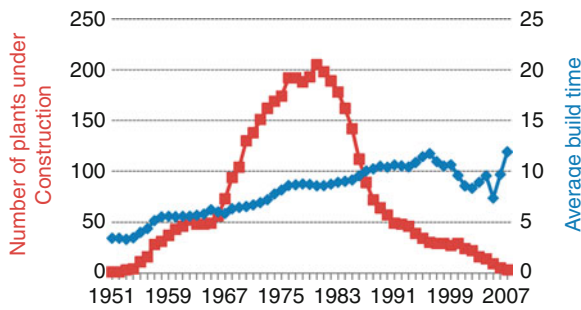
In general, Eq. 7 is integrable as long as $p(t)$ is known. While the formulas above may seem at first glance to be complicated, they are far simpler and more compact than their discrete discounting alternatives.

Lead and Lag Times in Cost Calculations

A critical factor in using the above equations, and when performing economic analyses in general, is to know when costs occur relative to a specified date. As already pointed out, this is captured in what are referred to as “lead” and “lag” times, and the reference time is typically taken to be the date of the facilities’ first operation. Components that affect the cost of nuclear power are shown schematically in Fig. 1. In very general terms, the costs associated with nuclear power production can be broken into *capital costs* (which include the construction of physical infrastructure such as the reactor plant, spent fuel storage facilities, or facilities involved in producing nuclear fuel, etc.), *operating costs* (which include operations and maintenance costs, taxes, fuel costs, as well as spent nuclear fuel disposal costs), and *decommission costs* (which include dismantling of physical infrastructure and site remediation).

Power Facility Construction

Historically, this has been by far the dominant factor affecting the cost of nuclear power (e.g., [3, 4]). The lead time is highly variable, Fig. 3, but 8 years or more is typical for modern large-scale facilities. How this cost is distributed depends on how a facility is financed, Fig. 2f.



Nuclear Power, Economics of. Figure 3

Reactors under construction and build times. The total number of nuclear power facilities under construction and the time for facility completion (from start of construction to date of first operation) are shown (The data are from the International Atomic Energy Agency Data Center)

Power Facility O&M

Operations and maintenance costs for the facilities are usually computed on a yearly basis and can be approximated with a uniform distribution as in Fig. 1b. Large capital outlays for replacement of major components can often be assumed to be onetime costs as in Fig. 2a (e.g., [3, 4]).

Cooling Water

Availability of cooling water has become a potential constraint for nuclear power plants in some locations. A facility's water use depends on its efficiency, as well as how it is cooled (open loop, closed loop, evaporation pond, etc.). A typical range of water withdrawals for modern Rankine cycle power plants is 2,000–4,000 l/MWh(e), for closed loop systems, and 100,000–220,000 l/MWh(e) for open-loop systems. Water consumption rates are considerably smaller. In this context, water consumption describes water that is taken from a source (ground or surface water), used, and not returned to that source. A cooling tower consumes water through evaporation for example. Water withdrawal describes water that is taken from a source, used, and then returned to that source. Cost estimates for water vary widely, but \$25 per 1,000 m³ has been reported and can be based on consumption, or withdrawal, rates and will depend on local water markets and regulatory structures (e.g., [15]).

Uranium Mining

The extraction of uranium ore typically occurs 2 years prior to its use as fuel in a reactor. Uranium is mined as U₃O₈ and is typically sold in this form. Modeled as a onetime cost that would recur with the refueling schedule of a reactor.

Uranium Conversion

The U₃O₈ requires conversion into UF₆ if the uranium is to be enriched to a higher ²³⁵U content that occurs in natural uranium. Uranium must be enriched for use as fuel in most commercial light water reactors worldwide [16]. The lead time for conversion is typically 1.5 years. Modeled as a onetime cost that would recur with the refueling schedule of the reactor.

Enrichment

Enrichment typically occurs 1 year before fuel placement in the reactor. Modeled as a onetime cost that would recur with the refueling schedule of the reactor.

Fuel Fabrication

Most reactors worldwide use uranium dioxide fuel that is surrounded by a protective “cladding” and configured in assemblies that can be placed into a reactor. Fabrication usually takes place 0.5 years before fuel emplacement. Modeled as a onetime cost that would recur with the refueling schedule of the reactor.

Leasing or Buying Nuclear Fuel

Not all nuclear utilities buy their fuel. Instead, many lease it, and sometimes from a company or trust that has been created specifically for the purpose of doing this. The lease company covers all expenses for the fuel (mining, conversion, enrichment, fuel manufacture transportation and storage) until it is onsite to be loaded into the core. During this time, the utility pays nothing. Once the fuel begins producing power, the utility will pay the lease company a prorated amount that covers the lease company's expenses plus some degree of profit. In some situations, this arrangement is financially advantageous, though this depends on the accounting practices of the utility and possible constraints from regulatory agencies. Typically, the total

cost obligation is met when the fuel has been used to completion. The cost function for the leased fuel can be linearly decreasing, Fig. 2e, or of some other shape [8].

Interim Spent Fuel Storage

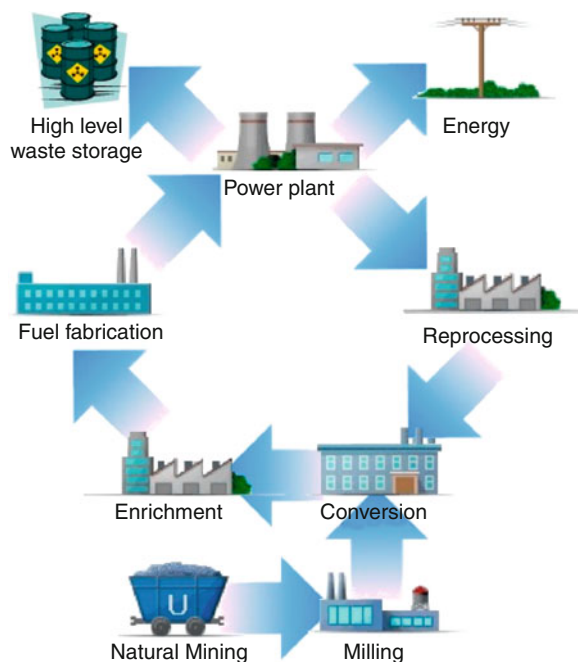
When the fuel reaches the end of its useful life, it is discharged from the core and is stored on site either under water or in air-cooled vaults until it is removed for final disposal or reprocessing. The residence time in interim storage is highly variable, and the spent fuel from some US reactors has remained in this type of onsite storage for decades. Spent fuel storage is typically calculated per unit time and per unit mass, and it is therefore a linearly increasing cost, Fig. 2e, at most reactor facilities, (e.g., [3, 4]).

Spent Fuel Recycle

Some countries, notably France, reprocess spent nuclear fuel and use the plutonium that it contains to manufacture what is called *mixed oxide fuel* (MOX – a mixture of plutonium oxide and uranium oxide) that can be used in addition to standard enriched uranium fuel, Fig. 4. In countries where this is done, spent fuel will cool in storage pools for 6–12 years before being sent to reprocessing. The distribution of this cost depends on whether funds are continuously set aside to meet this obligation or whether it is considered to be a series of onetime costs that occur as spent fuel is sent to recycle.

Spent Fuel Disposal

Most countries with domestic nuclear power assume that long-lived nuclear waste products will go into some form of geological storage. At present, only the US Waste Isolation Pilot Plant, in Carlsbad NM, is actively used for this purpose, and then only on a scale sufficient to handle transuranic waste from past US nuclear weapons efforts. As a result, the actual time at which spent fuel will leave interim storage for final disposal is unclear at most power plants (the final disposal of high-level waste from reprocessing is similarly unclear in countries where this is undertaken). Many reactor facilities whose spent fuel pools are nearing capacity have begun to transition fuel assemblies to onsite dry storage. In the United States,



Nuclear Power, Economics of. Figure 4

Recycle-based fuel cycle. A schematic representation of a reprocessing-based fuel cycle is shown. Instead of spent fuel going to storage, it is sent to reprocessing, and the plutonium, and possibly other *transuranic* elements (such as neptunium, americium, and curium) are recycled to make new fuel. The recycled transuranics would not, however, require enrichment. The high-level waste is comprised of byproducts of reprocessing (Figure courtesy of Robert Bell, The University of Texas at Austin)

the Nuclear Waste Policy Act of 1987 gives responsibility for the disposal of spent nuclear fuel to the Federal government [17]. In order to cover the associated expenses, US nuclear power facilities pay 1 mill/kWh(e) into the Nuclear Waste Fund for the purpose of covering the final disposal costs [17, 18]. The cost is incurred quarterly and, as a result, constitutes a series of recurrent one time costs for nuclear utilities in the United States, i.e., Fig. 2a.

Estimating Uncertainties

Equation 8 has the convenient feature of being linear with respect to the total cost of each fuel-cycle component. As a result, E_c can be scaled to account for changes in unit cost, provided that time points for the integral

in Eq. 7 remain fixed. Uncertainty in unit costs causes a corresponding uncertainty in the prediction of E_c . These effects can be accounted for by using the well-known formula for error propagation, where the variance of $E_c(x_i)$ is given by:

$$\begin{aligned} \text{var}(E_c) = & \sum_i \left(\frac{\partial E_c}{\partial x_i} \right)^2 \text{var}(x_i) \\ & + 2 \sum_i \sum_{j \neq i} \left(\frac{\partial E_c}{\partial x_i} \right) \left(\frac{\partial E_c}{\partial x_j} \right) \sqrt{\text{var}(x_i)} \sqrt{\text{var}(x_j)} r(x_i, x_j) \end{aligned} \quad (20)$$

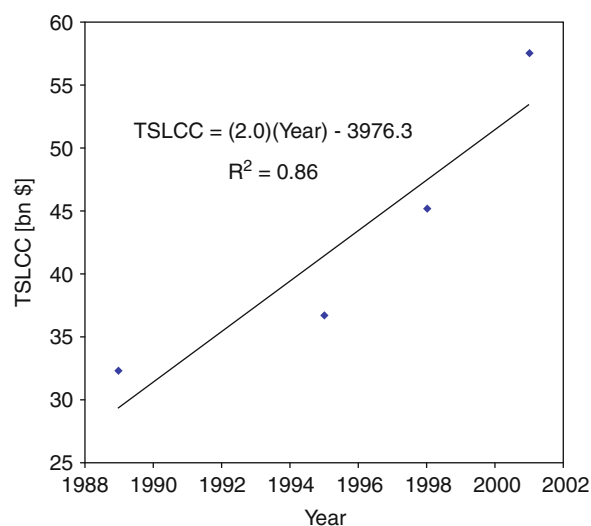
Here the inputs, x_i , represent the PV_i in Eq. 8 with respective variances $\text{var}(x_i)$. The term $r(x_i, x_j)$ is the correlation coefficient, 1 for fully correlated, -1 for anticorrelated and 0 for uncorrelated. The maximum and minimum variances are given by assuming that $r = 1, -1$ respectively with uncorrelated, $r = 0$, typically giving a variance that falls into the midrange. Equation 20 is much simplified in the case of $r = 0$. The electricity cost is assumed to have a Gaussian distribution which can be justified by the Central Limit Theorem [19], with the standard deviation of E_c being the square root of the variance. It should also be pointed out that Eq. 20 can be used for cost-sensitivity studies.

Costs and Their Uncertainties

Numerous studies have investigated the economics of nuclear power, notably the series of reports produced between 1987 and 2002 by the Organization for Economic Cooperation and Development and the Nuclear Energy Agency (OECD/NEA) [1, 13, 20–22] along with recent reports from groups at MIT and the University of Chicago. The 1994 OECD/NEA study [1], in particular, developed a framework for assessing the economics of nuclear fuel cycles. The study derived the expected levelized cost of a fuel cycle over the lifetime of a reactor, including transients (at startup and shutdown). The fuel cycle was divided into front-end components (uranium ore requirements, conversion to UF_6 , enrichment, fuel fabrication and transport) and back-end components (spent fuel transport, reprocessing, direct disposal, or high-level waste (HLW) vitrification and disposal). Cash outflows to meet these obligations were discounted to a reference date using a discrete model as was revenue from

electricity, and the subsequent expected cost in \$/kWh (e) was calculated.

The 1994 study [1] gave the most comprehensive cost estimates available at the time. Data were obtained from the literature and through survey of OECD member states and gave reliable results where industries were well established. Because no actual disposal facility existed for SF or vitrified HLW, these estimates were considered to be particularly uncertain [1]. The cost data were updated in the 2002 OECD/NEA study and estimated standard deviations were added [13]. Because no permanent repository is in operation, it is difficult to estimate what it would cost. However, the experience and cost studies at Yucca Mountain in the United States provides some indication. Here the cost estimates for the repository rose in constant dollars from \$32.2 billion to \$57.5 billion between 1989 and 2001, Fig. 5 [18, 23–26], for a repository that was designed to



Nuclear Power, Economics of. Figure 5

Yucca Mountain Life Cycle Cost over time. The Total System Life Cycle Cost (TSLCC) predicted for Yucca Mountain is shown over the period between 1985 and 2001 in year 2000 dollars (inflation adjustment done using GDP deflator). In that time, the predicted cost has nearly doubled from \$32.2 to \$57.5 billion in constant dollars, with the repository intended to hold 86,000 metric tons of spent fuel heavy metal [18, 23–26]. In this context, “heavy metal” refers only to the mass of uranium and transuranics within the spent uranium dioxide fuel

hold 86,000 metric tons of spent nuclear fuel. Representative costs, lead/lag times, and standard deviations for nuclear power systems are given in Table 2.

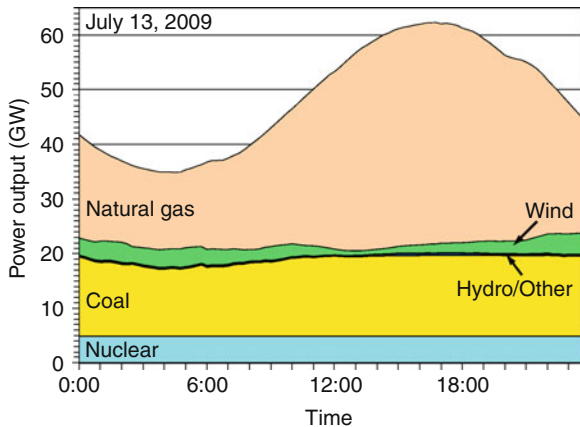
Discount and Interest Rates

The values of ρ that are used in discounting will have a significant effect on an economic analysis. As a result, discount rates have been widely discussed in the

context of decision theory, and in cost-benefit analyses of things that can have intergenerational effects such as environmental damage, resource allocation, or nuclear waste disposal. For economic comparisons, or cost studies, discounting accounts for the fact that payments made could instead have been invested and earned a rate of return. Alternately, future payments could be met by setting aside a smaller amount of money today and letting compound interest make up

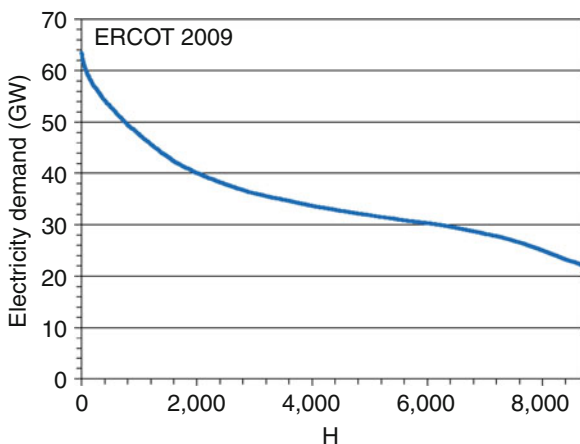
Nuclear Power, Economics of. Table 2 *Unit costs of various components of the fuel cycle for a light water reactor with and without spent fuel recycle.* For the variables listed U denotes uranium; Pu – plutonium; HM – heavy metal; SD – standard deviation; SW – separative work; % loss – the percentage of material lost during a particular fuel cycle step; SF – spent fuel; MOX – mixed oxide fuel; VHLW – vitrified high-level waste; and led/lag refers to the time at which a particular cost is incurred relative to loading of a fuel batch into the reactor. The lead time for construction is relative to the first startup of the reactor. References are given for each data column in order from *left to right*, NA designates “not applicable.” All costs are adjusted to year 2000 dollars. The standard deviation for the reactor facility construction cost was calculated from the cost data for seven reactors whose capital costs were given in [3]. It should be noted that many of the above costs are highly speculative, and other estimates are given in other studies, see for example [31]. Lag times for spent fuel recycle were estimated by assuming that SF_{UOX} remained in storage for 5 years and that fuel reprocessing occurred 1.5 years prior to MOX placement in the reactor. All other times are obtained by applying the time line for UOX fuel. The data in Fig. 4 suggests that the standard deviation for SF disposal is larger than that given in [13]

Fuel cycle component	Basis	\$	SD (%)	% loss	lead/lag (years)	References
Reactor facility construction	\$/kW(e) installed	3,560	15		-12 > -8	[11], [13], NA, see text
Reactor operations and maintenance	\$/kW installed/year	142	15			[13], same as capital costs, NA, NA
Uranium mining and milling	\$/kgU	30	33		-2.0	[13], [13], NA, [1]
U ₃ O ₈ to UF ₆ conversion	\$/kgU	5	40	0.5	-1.5	[13], [13], [29], [1]
Enrichment	\$/kgSW	80	37.5	0.5	-1.0	[13], [13], [29], [1]
UOX fabrication	\$/kgU	250	20	0.5	-0.5	[13], [13], [29], [1]
Onsite interim SF _{UOX} storage	\$/kgHM/year	30	5		+5	[13], [13], NA, NA
SF _{UOX} conditioning and disposal	\$/kgHM	668	24		+40	[13], [13], NA, [1]
Spent fuel recycle						
SF _{UOX} transport to reprocessing	\$/kgHM	50	20		+10.5	[13], [13], NA, see caption
SF _{UOX} reprocessing	\$/kgHM	800	12.5	2.0	+11	[13], [13], [13], see caption
MOX Fabrication	\$/kgHM	1,100	18	1.0	+12.5	[13], [13], [30], see caption
Onsite interim SF _{MOX} storage	\$/kgHM/year	30	5		+17.5	[13],[13], NA, NA
SF _{MOX} transport to disposal site	\$/kgHM/	50	20		+22.5	[13], [13], NA, see caption
SF _{MOX} conditioning and disposal	\$/kgHM	668	24		+57.5	Same as UOX, [13], NA, see caption
VHLW conditioning and disposal	\$/kgHM	288	12.5		+56	[13], [13], NA, see caption



Nuclear Power, Economics of. Figure 6

Electricity load curve and generation source. The figure shows the variation in electricity demand over a 24 h period, and the sources used to meet that demand (The data are from the Electricity Reliability Council of Texas, for July 13, 2009) (Figure courtesy of Stuart Cohen, The University of Texas at Austin)



Nuclear Power, Economics of. Figure 7

Load duration curve. The figure shows the number of hours during 2009 during which a given electric demand needed to be met. The base load for the year is shown at the far right (The data are from the Electricity Reliability Council of Texas, for July 13, 2009) (Figure courtesy of Stuart Cohen, The University of Texas at Austin)

the difference. In both cases, the appropriate discount rate would be one that reflects what one could obtain by investing the funds in bonds, stocks, or other investments, and receiving the market rate of return.

There is no single discount rate that is accepted for near-term analysis of utilities, nuclear power included, but values between 5% and 10% (in real terms) are common and reflect the historical range of return by the US utility sector (e.g., [4]). However, there are problems in using this discount rate for projects that run on intergeneration time scales. The dominant concerns are well summarized in several different studies (e.g., [27, 28]), but the main point is that a discount rate that is this high would suggest that money so invested would eventually grow to be larger than the domestic product of any country where the investment was made. As a result, some economists suggest that discount rates appropriate for intergenerational projects would have to tend toward a country's real rate of GDP growth and can be assumed to be around 1–2% (e.g., [27, 28]).

The question of which discount rate to use in evaluating nuclear power facilities is far from academic. Cost studies are often used to compare different systems or management strategies. A high discount rate can make the present value of far-off costs appear negligible. In other words, using a high discount rate in a costs-benefit study of a politically divisive issue, such as spent fuel disposal, can have the perverse effect of suggesting that it would always be cheaper to delay action. In fact, this type of argument has been advanced in the United States to advocate for pushing off the development of reprocessing or permanent geological disposal.

Cost Comparisons and External Costs

Using Eq. 8, or similar, one can calculate the cost associated with the generating electricity by various means. Results of such studies are shown in Tables 1 and 2. However, making true comparisons for the “cost” of generating electricity between different types of power systems can be complicated by the effect of externalities that are often difficult to capture in an economic analysis. Factoring in the true cost of carbon dioxide emissions would be an example. Carbon dioxide emissions from fossil-fueled plants are obvious. They are not so obvious for nuclear, wind, or solar power facilities. While nuclear power facilities have no direct emissions, there are carbon signatures associated with the materials out of which they are made,

the processes involved in facility construction, as well as for operations, maintenance, and decommissioning. The same is true for wind and solar power systems. Where one draws the boundary for an analysis (i.e., does one include the carbon dioxide needed to produce, say, the machinery for the processing of facility materials) can have a significant effect on the total carbon signature that is associated with a specific power source, and through this, its total potential cost. Because of the complexity of handling externalities (and the cost of carbon emissions is only one example), how best to do this remains an area of active research and debate.

Future Directions

Electricity demand varies with location, time of day, and time of year. The minimum demand at any point within a year is referred to as *base load*. The utility industry typically gauges demand requirements with what are called *load demand* and *load duration curves*, Figs. 6 and 7. Commercial nuclear power plants have historically been designed to operate at their maximum licensed power. The fraction of time that these facilities are in operation has also increased steadily over the years to an industry average of around 90%. As a result, commercial nuclear power is used to meet base-load requirements, along with coal, hydroelectric, and to a smaller extent wind. Additional load requirements are met with smaller-scale facilities (called *peaking power plants*) that typically run on oil or gas.

Because of perceived economies of scale, nuclear power plants have been built on an increasingly large scale. Recently though, this idea has begun to be revisited, and there are calls for the development of smaller-scale facilities with peak power outputs of as little as a few tens of megawatts. Such plants, known as *Small Modular Reactors*, could be used to meet local base-load requirements or as peaking plants. A shift in the industry to include smaller-scale facilities could have dramatic effects on the economics of nuclear power. Smaller facilities would likely have shorter build times and far lower capital costs. Many utilities also charge different rates, depending on when electricity is used, and this too would affect the economics of nuclear package plants that might be used to meet peak load requirements. (The first-generation light

water reactors were designed for power outputs of 300–600 MWe. In the United States, today most nuclear power facilities are rated for 800–1,000 MWe, but it is expected that the next generation of reactor will produce between 1,100 and 1,400 MWe.)

Acknowledgments

Special thanks to Andrew Osborne, Brady Stoll, and Geoff Recktenwald for editorial comments. Thanks to Eric Bickel and Michael Webber for useful discussions and to Nick Tsoulfanidis for editorial comments and for acting as Editor for this series.

Bibliography

1. OECD/NEA (1994) The economics of the nuclear fuel cycle. Technical report, OECD/NEA, Paris
2. Bunn M, Fetter S, Holdren JP, van der Zwaan B (2003) The economics of reprocessing vs. direct disposal of nuclear fuel. Technical report DE-FG26-99FT4028, Harvard University, Boston
3. Anolabehere S, Deutch J, Driscoll J, Holdren PE, Joskow PL, Lester RK, Moniz EJ, Todreas NE, Beckjord ES (2003) The future of nuclear power. MIT Press, Boston (Technical report)
4. Tolley GS, Jones DW (2004) The economic future of nuclear power. Technical report 2004, University of Chicago, Chicago
5. Garwin RL, Charkap G (2002) Megawatts and megatons – the future of nuclear power and nuclear weapons. The University of Chicago Press, Chicago
6. Goldschmidt P (2010) Multilateral nuclear fuel supply guarantees & spent fuel management: what are the priorities? *Daedalus* 139:7–19
7. Stern R (2007) The Iranian petroleum crisis and United States national security. *PNAS* 104:377–382
8. Cochran R, Tsoulfanidis N (1990) The nuclear fuel cycle: analysis and management, 1st edn. American Nuclear Society, La Grange Park
9. Brealey R, Myers S, Allen F (2002) Principles of corporate finance, 9th edn. McGraw-Hill Higher Education, New York
10. Schneider EA, Deinert MR, Cady B (2009) The cost impact of delaying the US spent nuclear fuel reprocessing facility. *Energy Econ* 31:627–634
11. Deutch J, Forsberg C, Kadak A, Kazimi M, Moniz E, Parsons J (2009) Update of the MIT 2003 future of nuclear power. MIT Press, Cambridge
12. White House (1993) Executive order #12866: regulatory planning and review. Washington, DC
13. OECD/NEA (2002) Accelerator driven systems and fast burner reactors in advanced nuclear fuel cycles. Technical report, OECD/NEA, Paris

14. Hannon B (1982) Energy discounting. *Technological Forecasting Soc Change* 21:281–300
15. Stillwell AS, King CW, Webber ME, Duncan IJ, Hardberger A (2011) The energy-water nexus in texas. (Special feature: the energy-water nexus: managing the links between energy and water for a sustainable future), *Ecol Soc* 16(1):2. <http://www.ecologyandsociety.org/vol16/iss1/art2/>
16. Benedict M, Pigford TH, Levi HW (1981) *Nuclear chemical engineering*, 2nd edn. McGraw-Hill, New York
17. Nuclear Waste Policy Amendments Act (1987)
18. DOE (2002) Nuclear waste fund fee adequacy: An assessment. DEO/RW-0534, DOE, Office of Civilian Radioactive Waste Management, Washington, DC
19. Shapiro SS (1990) Selection, fitting, and testing statistical models. In: Wadsworth HW (ed) *Handbook for statistical methods for scientists and engineers*. McGraw Hill, New York
20. OECD/NEA (1987) Nuclear Energy and its fuel cycle-prospects to 2025. Technical report, OECD/NEA, Paris
21. OECD/NEA (1993) The cost of high-level waste disposal in geological repositories – An analysis of factors affecting cost estimates. OECD/NEA, Paris
22. OECD/NEA (1996) Future financial liabilities of nuclear activities. Technical report, OECD/NEA, Paris
23. GAO (1990) Changes needed in DOE user-fee assessments to avoid funding shortfall. Technical report RCED-90-65, United States General Accounting Office, Washington, DC
24. DOE (1995) Analysis of the total system life cycle cost of the civilian radioactive waste management program. Technical report DOE/RW 0479, DOE, Office of the Civilian Radioactive Waste Management Program, Washington, DC
25. DOE (1998) Analysis of the total system life cycle cost for the civil radioactive waste management program. Technical report DOE/RW-0510, DOE, Office of Civilian Radioactive Waste Management Program, Washington, DC
26. DOE (2001) Analysis of the total system life cycle cost of the civilian radioactive waste management program. Technical report DOE/RW-0533, DOE, Office of Civilian Radioactive Waste Management, Washington, DC
27. Rabl A (1996) Discounting of long term costs: what would future generations prefer us to do? *Ecol Econ* 17: 137–145
28. Arrow K (1999) Discounting, morality, and gaming. In: Portney PR, Weyant JP (eds) *Discounting and intergenerational equity*. RFF Press, Washington, DC
29. Rasmussen N, Burke T, Choppin G, Croff G, Garrick J, Grunder H, Hebel L, Hunter T, Kazimi M, Kintner E et al (1996) *Nuclear wastes: technologies for separations and transmutation*. National Academy, Washington, DC
30. Shropshire D, Williams K, Boore W, Smith J, Dixon B, Dunzik-Gougar M, Adams R, Gombert D (2007) Advanced fuel cycle cost basis. US DOE, The Idaho national laboratory INL/EXT-07-12107

Nuclear Reactor Materials and Fuels

JAMES S. TULENKO

Laboratory for Development of Advanced Nuclear Fuels and Materials, University of Florida, Gainesville, FL, USA

Article Outline

Glossary

Definition of the Subject: Nuclear Reactor Materials and Fuels

Introduction to Nuclear Reactor Materials and Fuels
Nuclear Fuel Materials

Nuclear Clad Materials

Nuclear Moderator Materials

Nuclear Coolant Materials

Nuclear (Neutron) Poison Materials

Nuclear Structural Materials

Future Directions

Bibliography

Glossary

Austenitic stainless steel Austenitic steels contain alloys of chromium and nickel (sometimes manganese and nitrogen), structured around the Type 302 stainless steel composition of iron, 18% chromium, and 8% nickel. Austenitic steels are not hardenable by heat treatment. The most common austenitic stainless steel is type 304.

Burnup A measurement of the energy generated by fuel atoms that undergo fission. It is normally quoted in megawatt-days per metric ton of uranium metal or its equivalent (MWd/MTU).

Core plate In a reactor the upper and lower core plates supports the fuel, channels the cooling water into the fuel bundle, and assures each fuel bundle is maintained equidistant from each other.

Fertile fuel A material capable of creating a fissile fuel upon capture of a neutron. Examples are U^{238} and Th^{232} , which create Pu^{239} and U^{233} respectively.

Fissile fuel Capable of undergoing fission by thermal neutrons. The four primary nuclides are U^{233} , U^{235} , Pu^{239} , and Pu^{241} .

Fissionable fuel A material capable of undergoing fission, via the absorption of a neutron with kinetic

energy above a certain level. Examples are U^{238} and Th^{232} .

Half-life The time it takes for the mass of a substance undergoing decay to decrease by half.

Inconel Refers to a family of austenitic nickel-chromium-based high performance alloys trademarked by the Special Metals Corporation. Inconel metal is typically used in high temperature applications and is generally known for its resistance to oxidation and a superior ability to maintain integrity in high temperature conditions.

Isotopes Isotopes are different forms of atoms (nuclides) of the same chemical element, each having a different number of neutrons. Isotopes differ in mass number (the number of neutrons plus protons in the nucleus) but not in atomic number (total number of protons). All isotopes of an element have the same chemical properties, but frequently they have very different nuclear properties.

Martensitic or ferritic steels The major alloying addition in martensitic stainless steels is chromium in the range of 11–17%. The carbon levels can vary from 0.10% to 0.65% in these alloys. The high carbon enables the material to be hardened by heating to a high temperature, followed by rapid cooling (quenching). Martensitic steels offer a good combination of corrosion resistance and superior mechanical properties, offering maximum hardness, strength, and resistance to abrasion and erosion.

Neutron capture cross section The neutron cross section of an isotope is a measure of the probability of neutron capture by that element. It is the effective area that a particular atom of that isotope presents to absorb a neutron, and is measured by a unit (of area) called “barn,” which is 10^{-24} cm^2 .

Nuclear fuel clad or cladding The outer layer of the fuel rods, a barrier between the coolant and the nuclear fuel. It is made of a corrosion-resistant metal that has a low absorption cross section for thermal neutrons. In today’s modern commercial reactors, cladding is made usually of Zircaloy.

Passivating layer In the context of corrosion, passivation is the formation of a nonreactive surface film which inhibits further corrosion.

Plutonium (Pu) Man-made element with atomic number 94. Of the many Pu isotopes, the most noteworthy is ^{239}Pu because it is a fissile isotope and

may also be used to make a nuclear weapon. ^{238}Pu has found a great application in space missions powering electronic equipment.

Thermal conductivity Property of a material that describes its ability to conduct heat.

Thorium (Th) Naturally occurring element with atomic number 90; the only isotope found in nature is ^{232}Th . Th is more abundant than U. Upon capture of a neutron and subsequent radioactive decay, it forms ^{233}U , which is a fissile isotope.

Uranium (U) Naturally occurring element with atomic number 92; three isotopes are found in nature: U^{238} , U^{235} , and U^{234} with abundances respectively, of 99.28%, 0.711%, and 0.00057%. The isotope ^{235}U is the only naturally occurring fissile isotope.

Uranium dioxide (UO_2) An oxide of uranium; a black mildly radioactive, crystalline material that is used in nuclear fuel rods in nuclear reactors.

Zircaloy Is a group of alloys, mostly consisting of zirconium, with minor additions of tin (often about 1.5%), iron and chromium, used as a fuel rod clad material in light water thermal reactors.

Definition of the Subject: Nuclear Reactor Materials and Fuels

Nuclear reactor materials and fuels can be classified into six categories:

- Nuclear fuel materials
- Nuclear clad materials
- Nuclear coolant materials
- Nuclear poison materials
- Nuclear moderator materials
- Nuclear structural materials

The materials unique to nuclear reactors are the nuclear fuel materials. All the other materials are used in other applications. Because nuclear fuel materials are radioactive, they require specialized handling. Additionally, because of the connection of nuclear energy with the atom bomb, security in its tightest form is required for nuclear fuel.

Introduction to Nuclear Reactor Materials and Fuels

The first nuclear reactor was the famous “Chicago Pile” (CP-1) which was built under the west stands of Stagg

Field at the University of Chicago on a squash court and went critical on December 2, 1942. The first nuclear materials were uranium in both a metal and an oxide form as the fuel, carbon in a graphite form as the moderator, atmospheric air as the coolant (since the reactor power was essentially zero), cadmium as the control material, and wood as the structural material. Nuclear power plant designs have come a long way since then, particularly with regard to cooling. However, the fuel form remains basically the same.

Nuclear Fuel Materials

There are three basic nuclear fuel materials which can be utilized in many different forms: uranium, plutonium, and thorium. The most utilized fuel material is uranium and it is most often utilized in the oxide form in pellet form (see Fig. 1). That one pellet held in the tweezers has the energy equivalent of 1 t of coal. Natural uranium is composed of uranium 238 (99.3%) and uranium 235 (0.7%). Uranium 238 is classified as a fertile material, which produces the fissile Plutonium 239 when it captures a neutron. It is necessary to enrich the uranium in the uranium 235 isotope in order to run the light water reactors (LWRs). The uranium for commercial power reactors is normally enriched to ~5% U^{235} . The methods used to enrich the uranium are the gaseous diffusion process in which the isotopes of different molecular weights diffuse differently through a porous barrier and the gaseous centrifuge process, where the different masses are centrifuged in a fast rotating cylinder with the heavier molecules moving

toward the periphery of the centrifuge, thus separated from the lighter ones. For both processes a gaseous compound is necessary, that compound is uranium hexafluoride (UF_6) which is a solid at room temperature but becomes a gas at $56.4^\circ C$. As a gas, uranium hexafluoride becomes very suitable for a variety of processes to enrich the material in uranium 235. Lastly, and still in development, the laser method where laser radiation is used to excite, ionize, or dissociate one isotopic species without affecting the other may be used for enrichment. The current favored method of enrichment is the gaseous centrifuge.

Uranium as a fuel is also utilized in the metallic form as an alloy with either zirconium or aluminum in Sodium Fast Reactors (SFRs) and in research reactors. Experimental developmental work is researching the carbide and nitride forms of uranium as a useful fuel form for SFR reactors and Gas Cooled Reactors (GCRs) because of their excellent thermal conductivity properties and higher uranium density. Research is also under way to utilizing uranium or plutonium in either the oxide or metal form with materials such as silicon carbide or zirconium to form inert matrix fuel forms with improved thermal conductivity properties. Uranium dioxide (UO_2) is a good fuel because it has a high melting point ($2,865^\circ C$) and does not dissolve in water. However, in an oxidizing environment UO_2 can be oxidized to U_3O_8 , which does dissolve in water. The major drawback of uranium dioxide as a fuel form is that it has a relatively low thermal conductivity, which continues to decrease with temperature until approximately $2,000^\circ C$. Additionally, the thermal conductivity of uranium also decreases with burnup as the fuel produces power. Since the Nuclear Regulatory Commission does not permit reactors to operate in a condition where centerline melting of the fuel pellet may occur, the high melting temperature of the uranium dioxide does tend to offset its low thermal conductivity. However, the high temperatures in the uranium oxide pellet during operation does lead to thermal stresses, which result in cracking in the pellet both in the radial and circumferential directions. The uranium carbide (UC) and uranium nitride (UN) fuel forms have the benefits of a very high thermal conductivity and a higher density than UO_2 of uranium. However, they dissolve in water and therefore cannot be considered for use in the water cooled



Nuclear Reactor Materials and Fuels. Figure 1
Uranium dioxide (UO_2) fuel pellet (Courtesy AREVA)

Nuclear Reactor Materials and Fuels. Table 1 Comparison between UO_2 , UZr, UC, and UN fuel for key material characteristics

Fuel comparison	Uranium oxide (UO_2)	Metallic fuel (UZr)	Carbide fuel (UC)	Nitride fuel (UN)
Density (g/cm^3)	9.75	14	13.58	13.53
Melting point ($^\circ\text{C}$)	2,750	1,080	2,420	2,780
Thermal conductivity (W/mK)	2.9	14	16.5	14.3

reactors. The metal fuel form has a higher thermal conductivity and a higher density of uranium than UO_2 , but does have a lower melting temperature. The material properties of the various uranium fuel forms are shown in Table 1.

Plutonium does not exist in nature, but is a totally man-generated fuel. Plutonium 239 is a fissile isotope generated from a neutron being captured in U^{238} and has a larger fission cross section than that of uranium 235. However, plutonium 239 has a smaller half-life of 24,200 years versus the half-life of uranium 235 of 7.04×10^8 years, so it is far more radioactive and cannot be handled like uranium. Additionally, plutonium 239 in reactor forms a whole family of fertile and fissionable isotopes: Pu^{240} , Pu^{241} , and Pu^{242} . These isotopes have similar short half-lives of 6,537 years for Pu^{240} , 14.4 years for Pu^{241} , and a long half-life of 376,000 years for Pu^{242} . Plutonium metal melts at approximately 640°C , a much lower temperature than uranium, and has a thermal conductivity at 300 K of 6.74 W/mK versus a value of 27.5 W/mK for uranium metal. Plutonium fuel is normally used as an oxide and is mixed with natural or depleted uranium oxide. This mixed oxide fuel form is termed MOX.

Thorium is a fertile fuel containing only the single thorium isotope Th^{232} , which upon capture of a neutron forms Th^{233} . Th^{233} is transmuted by beta decay with a half-life of 22.3 min into Pa^{233} , which beta decays with a half-life of 27 days into the fissile uranium isotope U^{233} . Thorium dioxide's advantage over uranium dioxide as a fuel form is that it has both a higher thermal conductivity

($\sim 5.25 \text{ Wm}^{-1} \text{ K}^{-1}$ at 800 K), a higher melting point ($3,350^\circ\text{C}$), 80°C higher than uranium, and is insoluble in water under all conditions, having only the single oxidation state (O_2). Thorium oxide with enriched uranium oxide was the fuel in the very first full-scale nuclear power reactor (the Shippingport Atomic Power Station), which was built by Admiral Rickover for the Duquesne Light Company. This reactor went critical on December 12, 1957.

Nuclear Clad Materials

The clad materials, which enclose the nuclear fuel, are really the first line of defense to retain the radioactive materials within its boundaries and to protect the fuel from the moderator. In addition to these tasks, the clad also provides structural integrity for the fuel assembly and a surface for heat transfer from the fuel to the coolant. The fuel clad and the guide tubes, which connect the upper and lower end fittings of the fuel assembly, are normally made of the same material. The cladding material must be chemically compatible with the fuel and it must be corrosion resistant with the coolant material. Additionally, it is important that the cladding and the guide tube material have a very small neutron capture cross section so as to not interfere with the fission process. Stainless steel has excellent corrosion properties and an excellent resistance to neutron irradiation. However, it has a relatively high thermal neutron absorption cross section. Until a material could be developed which had a low neutron absorption cross section and good corrosion and structural properties under irradiation, stainless steel was used as both the cladding material and the guide tube for both thermal and fast reactors. However, in thermal reactors, the stainless steel neutron absorption cross section is relatively large and the resulting penalty in neutron economy forced a search for a replacement material which would have the required mechanical strength, corrosion resistance in the water moderator, and a low neutron absorption cross section. This search led to the development of zircaloy, a zirconium alloy material made with minor additions of tin, oxygen, iron, and chromium. The zircaloy family of materials was developed out of the US Naval Reactors program and was quickly picked up by the commercial industry to replace stainless steel as a cladding material and guide

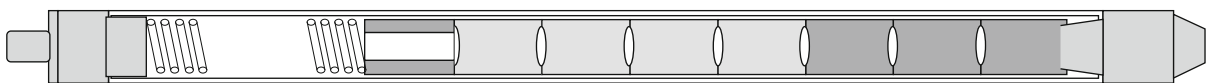
tube material (and shroud for BWR reactors) both for pressurized and boiling water reactors. The Shippingport Reactor, mentioned earlier, utilized zircaloy clad for its fuel rods. It should be noted that the light water reactor (LWR) concept was developed for the nuclear navy by Admiral H. Rickover and that he helped to shape the nuclear industry. Zircaloy had the required mechanical strength, a small neutron cross section, and good corrosion resistance. While zircaloy reacts with oxygen, it does form a passivating layer that protects the remaining zircaloy from further oxidation. Recent zircaloy alloy developments have added niobium and eliminated the chromium to give better oxidizing and hydriding resistance. Zircaloy is still the material of choice for the nuclear commercial industry. The upper and lower end fittings are made either of stainless steel or of inconel, since they are outside of the active nuclear portion of the core. Different inconels have widely varying compositions, but all are predominantly nickel, with chromium as the second element and iron as the third element. Nickel has a large thermal capture cross section for neutrons. The spacer grids which hold the fuel rods in place were originally either inconel or stainless steel, but were subsequently also changed to zircaloy. Figure 2 displays a schematic of a nuclear fuel rod showing the fuel pellets, top insulator pellet, and the fuel hold down spring which keeps the fuel pellets from moving during shipping and handling. Figure 3 shows a fuel assembly structure with end fittings, guide tubes, and spacer grids, but no fuel rods. The fuel guide tubes provide the vertical structure for the fuel assemblies and serve as channels for the control rods to enter into the fuel assemblies. The spacer grids provide the horizontal structure for the fuel assemblies. Figure 4 shows a fuel assembly with fuel rods in place.

Stainless steel continues to be used as the cladding for fast reactors, since stainless steel tends to have low capture cross sections in the fast neutron energy region.

However, even here, as coolant temperatures have been increasing in reactor designs in order to obtain ever higher thermal efficiencies, improved steel alloys have been required to be developed. First, austenitic stainless steels replaced the 304 stainless, and now oxide dispersion strengthened (ODS) martensitic or ferritic steels are under development to replace the austenitic stainless steels. Even more recently, silicon carbide cladding has been being researched as a cladding for the future. The silicon carbide cladding is made from a silicon carbide weave which is then impregnated with a silicon carbide vapor and in the latest designs a metal liner is placed on the inner surface of the clad. As a ceramic, silicon carbide would be resistant to oxidation and would have an extremely high melting temperature, so as to render the loss of coolant accident (LOCA) a nonevent.

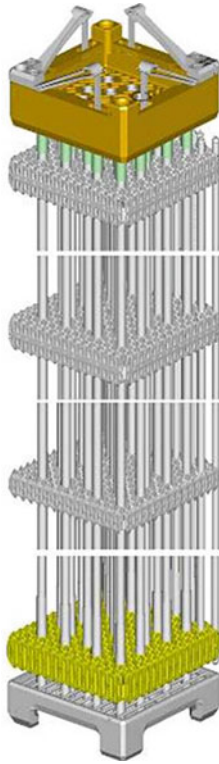
Nuclear Moderator Materials

With the advent of nuclear reactors, the nuclear fission process was found to proceed more easily if the neutrons were moderated in energy into the thermal range ($E < 1.0$ eV). The ideal moderator would have a very low thermal neutron capture cross section and a low atomic number in order to increase the neutrons energy loss on each collision. The very first nuclear reactors used graphite (carbon) to slow the fission neutrons down to the thermal energy regime. Carbon had a very small neutron absorption cross section and an atomic weight of 12. With the advent of the power reactors, water was used as both the moderating and the coolant material. Hydrogen, with an atomic weight of 1, equal to that of a neutron, could slow the neutrons down with very few collisions, and moreover, water is inexpensive. However, hydrogen does have a relatively high thermal neutron capture cross section, turning the hydrogen into deuterium. This capture cross section of ordinary water requires that light water reactors must use enriched uranium as a fuel. Deuterium oxide, or heavy water,



Nuclear Reactor Materials and Fuels. Figure 2

Schematic representation of a fuel rod showing fuel pellets, upper fuel spacer, fuel clad, fuel hold down spring, and upper and lower fuel end plugs



Nuclear Reactor Materials and Fuels. Figure 3
Pressurized water nuclear fuel assembly structure showing upper and lower end fittings, fuel guide tubes, and fuel spacer grids – without fuel rods in place (Courtesy AREVA)

would be an ideal moderator and coolant because of its very low capture cross section for neutrons. However, its very high price makes it a very expensive material to use. The Canadians utilize heavy water as their moderator and coolant in their Canadian-Deuterium-Uranium (CANDU) reactors. The extremely low absorption cross section of deuterium allows the CANDU to construct and operate nuclear power plants without the need for enrichment, utilizing natural uranium fuel. But they had to build plants to separate the heavy water (deuterium) out of ordinary water. Hydrogen has also been used in solid form in zirconium hydride (ZrH) as a moderator for space applications and for the TRIGA (Training, Research, Isotopes, General Atomics) research reactors of General Atomics. Lithium, a liquid metal, can also be used as a moderator. Beryllium has been used as a moderator in research reactors and in many ways is superior to graphite and lithium. However, beryllium is a very toxic material



Nuclear Reactor Materials and Fuels. Figure 4
Shows a fuel assembly with fuel rods in place (Courtesy AREVA)

which limits its usefulness because of Environmental Protection Agency restrictions on its use.

Nuclear Coolant Materials

Nuclear coolant materials should be marked by a low melting temperature and a high boiling point. The obvious coolant which has been used in fossil boilers for years is water or steam. However, its low boiling point requires that it be pressurized to allow for a high heat content. Thus, the pressurized water reactor (PWR) has a system pressure of approximately 2,200 psi and the Boiling Water Reactor (BWR) has a pressure of approximately 1,100 psi. As mentioned earlier, the Canadians utilize heavy water as their coolant in the CANDU reactor. However, the Canadians are developing new reactors which utilize ordinary water as the coolant, while retaining heavy water as the moderator. The British for their reactors went with graphite as

the moderator and pressurized carbon dioxide (CO_2) as the coolant in their Advanced Gas Reactor (AGR). The United States has researched using pressurized helium cooled, graphite moderated reactors at the Peach Bottom Reactor (in Pennsylvania) and the Fort Saint Vrain Reactor (in Colorado). The fast reactors utilize either sodium or helium as a coolant. Various experimental reactors have utilized exotic coolants such as fluoride salts, lithium, sodium-potassium alloy (NAK), and a lead-bismuth eutectic. Lead-bismuth (Pb-Bi) eutectic has a low melting point, a high boiling point, and does not react with water or oxygen. The Russians have operated lead cooled reactors in their nuclear navy. The heavy density of lead may generate erosion in its container pipes, particularly bends in its flow path. Molten salts have the advantage of having a very high boiling point and containing the fuel right in the molten salt mixture, thereby doing away with a need to fabricate the fuel.

Nuclear (Neutron) Poison Materials

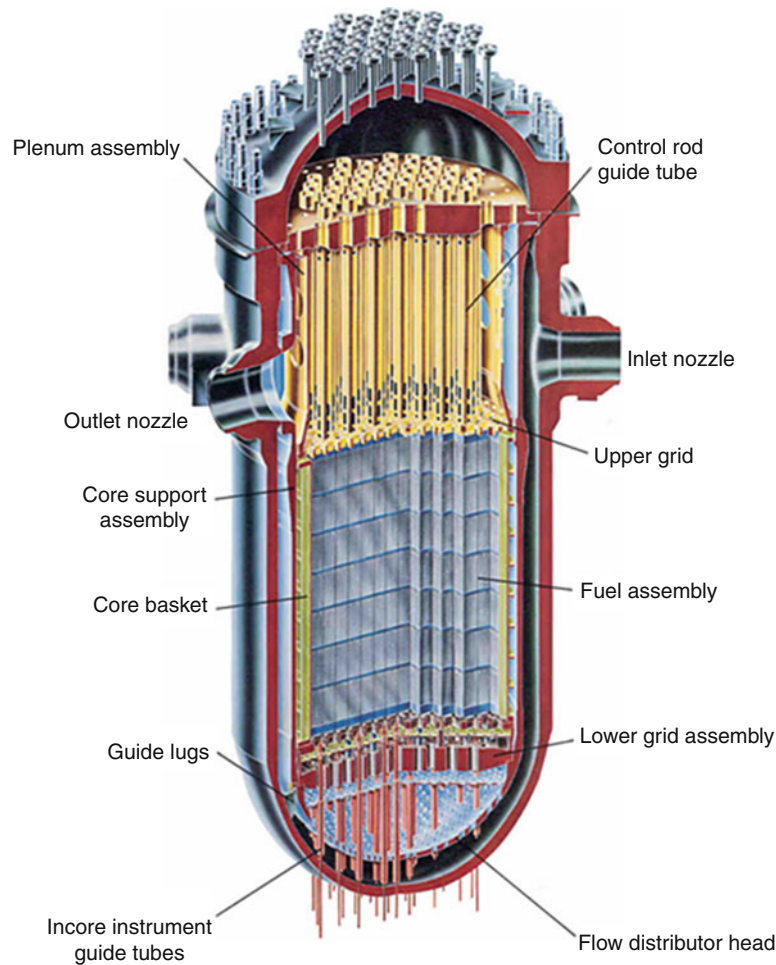
Nuclear engineers use the term “poison” to describe any material that steals neutrons away from the fission process. Nuclear reactors use poison materials to control the fission chain reaction. These poison materials must have a high neutron capture cross sections. Poison materials are of two types: burnable and nonburnable. In the burnable poison type, the poison material upon capturing a neutron, loses its large neutron capture cross section. In the nonburnable poison type, the poison materials, upon capturing a neutron, convert into a similar high-absorbing neutron cross section material. Three typical control materials used in control rods are (1) hafnium, (2) an alloy of silver, indium, cadmium, and (3) boron. Boron is not a nonburnable poison, like the other control materials, because the main neutron-absorbing isotope ^{10}B , upon capturing a neutron, converts to ^4He and ^7Li . Boron control rods must be replaced at regular intervals, unlike control rods made with hafnium or the alloy of silver, indium, and cadmium. Soluble boron is added as a poison material into the coolant of pressurized water reactors, where it can be adjusted to keep the reactor in a critical condition. When used in the coolant, the boron is called a chemical shim or “chem shim.” The use of boron in the coolant allows pressurized water reactors to keep

the control rods out of the core during operation. In a PWR, control rods are used only to shut down the reactor. Boiling water reactors cannot use boron in the coolant because of the boiling action which would lead to boron plate outting and also that the boron could reach the turbines and ruin the blades. For these reasons, they must utilize control rods in the core during operation in order to control the reactivity of the core.

Boron is also used as a burnable poison in burnable poison rod assemblies (BPRAs), which are inserted into the unoccupied control rod guide tubes and changed at each refueling. The boron is depleted during the cycle, such that at end of cycle the poison effect has been exhausted. Boron is also placed inside the fuel rods in Westinghouse reactors where a thin layer of zirconium diboride (ZrB_2) is coated on the fuel pellets. Gadolinium and erbium are also used as burnable poison materials and are mixed with the uranium dioxide (UO_2) fuel when forming the fuel pellets. BWR reactors use gadolinium extensively in the fuel in order to control the reactivity because of their inability to utilize boron in the coolant. The advantage of the burnable poison in the fuel, as opposed to BPRAs, is that there is less nuclear waste and the displacement of the water in the guide tube by the BPRAs does result in an end-of-cycle reactivity penalty for the BPRAs.

Nuclear Structural Materials

As mentioned, the very first reactor (CP-1) used wood as the structural material. However, today the nuclear structural materials are steel and concrete. The nuclear-fueled core is contained inside a steel vessel which has steel core support plates, top and bottom, upon which the core is held in place. The steel vessel is approximately 6 in. thick. In new designs, the steel vessel may be inside a concrete shield wall. In pressurized water reactors, steel baffles surround the core and separate the incoming colder water in the down comer from the heated water in the open core. In the boiling water reactor, each fuel assembly is contained in a zircaloy fuel shroud that isolates the water being heated in each fuel assembly. In both the PWR and the BWR, the steel pressure vessel is contained inside a steel reinforced concrete containment structure with a steel liner. [Figure 5](#) shows a PWR reactor vessel with the fuel core, upper core internals, and upper and lower grid assemblies.



Nuclear Reactor Materials and Fuels. Figure 5

Pressurized water reactor vessel showing fuel core, upper and lower grid assembly, and core internals (Courtesy AREVA)

Future Directions

For light water reactors, the future directions for the thermal reactors include the development of a crushable fuel pellet design and a ceramic cladding, such as the silicon carbide clad mentioned. This crushable pellet is needed because ceramic clad materials do not do well in tension. Thus, it is necessary for the pellet to give when interacting with the cladding, to prevent creating excessive tension forces on the clad. This development would also eliminate concerns regarding the pellet clad interaction (PCI) which currently limits power-transitions during operation of light water reactors.

The successful development of the molten salt reactor would offer many benefits by eliminating the fuel fabrication step, allowing for online reprocessing and fueling,

and also allowing for higher temperature operation with a resultant increase in thermal efficiency. In the long run, future reactors will most likely be fast breeder reactors, either using sodium or helium as the coolant, and fueled with uranium carbide and clad with silicon carbide.

Bibliography

1. Cochran RG, Tsoulfanidis N (1999) The nuclear fuel cycle: analysis and management, 2nd edn. American Nuclear Society, La Grange Park. ISBN 0-89448-451-6
2. Knief R (1992) Nuclear engineering: theory and technology of commercial nuclear power, 2nd edn. Hemisphere, Washington, DC. ISBN 978-0-89448-458-2
3. Glasstone S, Sesonske A (1994) Nuclear reactor engineering: reactor design basics, vol 1. Chapman & Hall, New York. ISBN 0-442-20057-9

Nuclear Safeguards and Proliferation of Nuclear Weapons Materials

MICHAEL C. BAKER

Advanced Nuclear Technology Group, Nuclear Nonproliferation Division, Los Alamos National Laboratory, Los Alamos, NM, USA

Article Outline

Glossary

Definition of the Subject

Introduction: Historical Development of Nuclear Safeguards and Nonproliferation Policy

Technologies for Safeguarding Nuclear Materials

Detection of Nuclear Proliferation and Treaty Verification

Detection and Prevention of Illicit Trafficking and Terrorism

Future Directions

Bibliography

Glossary

Calorimetry Calorimetry is a nondestructive assay technique for determining the thermal power output of heat-producing nuclear materials. Calorimeter systems are used to determine the power output (Watts) of various radionuclides over a broad range of power levels and sample types (Pu, highly enriched uranium, and Tritium).

Destructive assay Destructive Assay (DA) aims at the measurement of the nuclear material content, the isotopic composition, or other chemical properties of a sample. The analysis introduces a significant change to the physical form of the test sample.

Detection time The maximum time that may elapse between diversion of a given amount of nuclear material and detection of that diversion by a safeguards authority.

Direct-use material Nuclear material that can be used for the manufacture of nuclear explosive devices without transmutation or further enrichment. It includes plutonium containing less than 80% ^{238}Pu , highly-enriched uranium, and ^{233}U .

Diversion The undeclared removal of nuclear material from a safeguarded facility or the use of a safeguarded facility for the production or processing of undeclared nuclear material.

Dual-use technology Technology that can be used for both military and peaceful purposes.

Highly enriched uranium (HEU) Uranium containing 20% or more of the isotope uranium-235.

Holdup The amount of nuclear material remaining in process equipment and facilities after the in-process material, stored materials, and product have been removed.

Nondestructive assay Nondestructive Assay (NDA) techniques allow analysis of a sample in its own physical/chemical form without any modification of its properties and characteristics. The most common NDA techniques applied in nuclear safeguards include gamma spectroscopy, neutron counting, and calorimetry.

Nuclear material (NM) Any source or special fissionable material.

Safeguards An integrated, layered system of physical protection, material accounting, and material control measures designed to deter, prevent, detect, and respond to unauthorized possession, use, or sabotage of nuclear materials.

Significant quantity (SQ) The approximate amount of nuclear material for which the possibility of manufacturing a nuclear explosive device cannot be excluded. The International Atomic Energy Agency uses 8 kg of plutonium, 25 kg of ^{235}U in highly enriched uranium, and 75 kg of ^{235}U in low enriched or natural uranium.

Source material Uranium containing the mixture of isotopes occurring in nature; uranium depleted in the isotope 235; thorium; or ore concentrate.

Special fissionable material Plutonium-239; uranium-233; uranium enriched in the isotopes 235 or 233; and any material containing one or more of the foregoing; and such other fissionable material as the International Atomic Energy Agency Board of Governors shall from time to time determine; but the term "special fissionable material" does not include source material.

Special nuclear material (SNM) Plutonium, uranium-233, uranium enriched in the isotope 235, and any other material which, pursuant to

Section 51 of the United States Atomic Energy Act of 1954, as amended, has been determined to be special nuclear material, but does not include source material; it also includes any material artificially enriched by any of the foregoing, not including source material.

Definition of the Subject

Nuclear technology, by its very nature, may be used for peaceful or harmful ends. From its birth, efforts have existed to control access to, as well as the spread of, nuclear technology for destructive purposes, while at the same time attempting to provide access to its multitude of beneficial applications. Nuclear safeguards refers to the system of legal, institutional, and technical mechanisms used to control the use of nuclear materials, detect nuclear material misuse, and if the material is misused, to provide domestic and international responses in a timely fashion. Proliferation refers to the spread of weapons usable or direct-use materials, nuclear weapons technology, and the knowledge and skills necessary for the design and manufacture of such weapons. Nonproliferation policy and technology, including nuclear safeguards, has traditionally been centered on the spread between nation states. In the past decade, that focus has broadened to include the spread of materials, technology, and knowledge to criminal and terrorist entities. Nuclear safeguards must be viewed as a multilayered system, no single layer being self-sufficient, which as a whole, provides deterrence, detection, and response to theft, diversion, or misuse of nuclear material and nuclear weapons knowledge and skills.

Introduction: Historical Development of Nuclear Safeguards and Nonproliferation Policy

Before the end of World War II, diplomats and scientists familiar with nuclear weapons technology and its early development during the war were struggling with how to control the technology while sharing its benefits. Early efforts were largely focused on international control and the development of an international authority for management of the control regime and nuclear materials. In June 1946, following the formation of the United Nations Atomic Energy Commission (UNAEC) in January, the US representative, Bernard Baruch, proposed a plan for such an international

control regime. The Soviet Union and other countries opposed the significant transfer of state authority to an international body, as would be required, and the Baruch plan quickly came to pass. However, it did provide the conceptual framework and foundation for subsequent discussions between nation states.

By mid 1953, three nations (USA, United Kingdom, and Soviet Union) had conducted nuclear tests, Dwight Eisenhower had been elected President of the USA, and Joseph Stalin had died. These events set the stage for President Eisenhower's address to the United Nations on December 8, 1953. This address outlined the President's Atoms for Peace initiative, which included the creation of an international agency responsible for the peaceful expansion of nuclear technology and the control of nuclear materials that in accordance with the President's plan would be contributed by nation states. The President's initiative also called for states receiving assistance with the peaceful application of nuclear technology to allow inspections that would ensure that the technology and materials were not used for military purposes. The President's speech was the genesis of the International Atomic Energy Agency (IAEA).

In 1956, 81 nations unanimously approved the charter of the IAEA. At the same time, international cooperation related to the peaceful application of nuclear technology was rapidly expanding. By the end of 1959, the USA had cooperative bilateral agreements with 42 countries for collaborative nuclear technology development. Within the next decade, the Soviet Union would have similar agreements with another 26 countries. In most cases, both the USA and Soviet Union included at least a minimal acknowledgement of the dual-use capabilities of the nuclear technologies to be shared. The Soviet cooperative agreements typically required the return of spent nuclear fuel to the Soviet Union as a nuclear material control measure, as well as a pledge to use the material only for peaceful purposes. Most of the US agreements foresaw the need for nuclear safeguards to be applied by the USA that would eventually be turned over to the IAEA. The safeguards experience gained as a result of these bilateral agreements, as well as that obtained following the establishment of Euratom safeguards by the European community in 1957, provided a useful experience base and context for the establishment of specific IAEA safeguards agreements and protocols. In 1959,

IAEA safeguards were applied for the first time at the Japanese JRR-3 research reactor running on natural uranium fuel from Canada. [1]

The political history of the early 1960s included France and China joining the club of countries that had conducted nuclear tests and the Cuban Missile Crisis in 1962. This later event, one of the major confrontations of the Cold War, is the closest the world has come to a nuclear war. During the crisis, US armed forces were at their highest state of readiness and Soviet commanders were prepared to use nuclear weapons to defend Cuba against an invasion [2]. As a result, both superpowers came to the realization that greater cooperation on nuclear arms and nuclear safeguards were necessary for their mutual national security and that of their allies. This recognition and associated efforts to seek common ground were also seen in other countries. For example, in 1967, the Treaty for the Prohibition of Nuclear Weapons in Latin America was signed. By 1968, negotiations regarding the Treaty on the Non-Proliferation of Nuclear Weapons (NPT) were well underway. Although the idea of such a treaty had been proposed ten years earlier by Ireland, the NPT was not opened for signature until 1970. This marked a major milestone in the evolution of nonproliferation policy and nuclear safeguards. One of the significant requirements that resulted from the NPT is that nonnuclear weapons states place all of their peaceful nuclear programs under IAEA safeguards. As of 2007, the NPT had been ratified by 189 nations.

After the NPT had established a legal framework for international safeguards in 1970, the next twenty years saw the rapid development of both destructive and nondestructive measurement technology in support of nuclear safeguards, as well as the continued development of procedures and protocols used for international inspections. Formal training programs were established in multiple countries, as well as by the IAEA, to train and develop a cadre of professional safeguards inspectors. Some would argue that safeguards programs, personnel, and international policy had reached a state of maturity by the mid 1980s. However, in 1991, the international community was shocked with the revelation of a clandestine nuclear weapons program in Iraq. This program had developed under the noses of international inspectors that had been reviewing the declared activities in Iraq but were unaware of the clandestine program.

The international community responded with a reevaluation of safeguards and nonproliferation policy that included an acknowledgment of the need for IAEA safeguards to be significantly strengthened. Specifically, the IAEA needed the capability to detect undeclared nuclear activities in countries that were parties to the NPT. In 1997, the IAEA Board of Governors approved the “Additional Protocol” which allows for the use of environmental sampling to detect clandestine activities, earlier provision of design information on nuclear facilities to the IAEA, complementary access to additional locations within a country, and a broader use of open source, satellite imagery, and third party information to access nonproliferation performance in a particular nation. As of 2009, implementation of the Additional Protocol was complete or at a minimum underway in countries with significant nuclear programs.

Until the terror attacks on September 11, 2001, the diversion of nuclear material for a nation state’s weapon program was the primary threat that nuclear safeguards programs were designed to mitigate and international policy experts had been focused. Those attacks demonstrated a higher level of sophistication than previously expected and brought the realization to international security professionals that nuclear terrorism sponsored by substate actors must also be considered as a significant threat to be addressed by nuclear security programs. In the last decade, the proliferation of nuclear materials and nuclear weapons to substate groups has replaced a potential nuclear confrontation between dominant superpowers as the preeminent global security concern. This has shifted emphasis in the nuclear security community to border monitoring, event attribution, and consequence management. In 2009, with the election of a new President in the USA, focus of international security policy and programs is returning to the threat from nations with large nuclear arsenals; however, it is certain that the terrorist threat will remain a greater concern than it was before 2001. Significant events that impacted the development of international nonproliferation and safeguards policy are shown in [Table 1](#).

Technologies for Safeguarding Nuclear Materials

To properly secure nuclear materials, a state must know how much material it possesses and where it is located. This baseline is used to plan protection strategies,

Nuclear Safeguards and Proliferation of Nuclear Weapons Materials. Table 1 Influential events in the development of international safeguards and nonproliferation policy

Date	Event
1945	USA's first nuclear test at the Trinity Site.
	Nuclear bombing of Hiroshima and Nagasaki.
1946	Formation of the United Nations Atomic Energy Commission (UNAEC).
	Baruch Plan presented.
1949	Soviet nuclear test.
	UNAEC dissolves.
1952	United Kingdom nuclear test.
1953	President Eisenhower presents Atoms for Peace.
1955	USA and Turkey conclude the first "Atoms for Peace" cooperative agreement for the peaceful uses of nuclear energy.
1956	International Atomic Energy Agency (IAEA) charter is approved.
1957	European Atomic Energy Community (Euratom) is established.
1959	First application of IAEA safeguards.
1960	French nuclear test.
1962	Cuban Missile Crisis.
1964	Chinese nuclear test.
1967	Treaty for the Prohibition of Nuclear Weapons in Latin America established.
1968	Treaty for the Non-Proliferation of Nuclear Weapons (NPT) established.
1974	Indian nuclear test.
1979	Three Mile Island nuclear accident.
1986	Chernobyl nuclear accident.
1991	Iraqi clandestine nuclear weapons program discovered.
1995	NPT is made permanent.
1997	IAEA approves the Additional Protocol.
1998	Pakistan's nuclear test.
2001	September 11, 2001 terrorist attacks in the USA.
2003	Disclosure of the A. Q. Kahn clandestine nuclear trade network.
2006	North Korean nuclear test.

which mitigate the risks from both insider and outsider threats. The fundamental system that both provides this baseline information and allows the tracking and maintenance of the information as it changes over time is the nuclear material accounting system. To be effective, this system must be supported by a nuclear material measurement program, which provides accurate and precise information on the material being protected.

Nuclear material accounting and nuclear material measurements must be closely coupled with a physical protection system that ensures that access to the material is controlled and the materials stay in their intended locations. The physical protection system consists of people, procedures, and technology, which are designed and implemented following a careful evaluation of system vulnerabilities against predicted threats. Each facility is unique; therefore, a systematic planning effort is required to manage risks appropriately.

Nuclear Material Accounting

Nuclear material accounting is not that different from managing inventory for a complicated manufacturing and warehouse operation or the accounting involved in today's banking systems. Matter of fact, initial nuclear materials accounting developed from procedures and tools, which had been developed for financial and materials management. Today, nuclear material accounting is defined as the use of statistical and accounting measures to maintain knowledge of the quantities of SNM present in each area of a facility. It includes the use of physical inventories and material balances to verify the presence of material or to detect the loss of material after it occurs.

The quality of nuclear material accounting data and therefore the quality of the entire safeguards system is dependent on

- A measurement system that provides accurate, repeatable data on nuclear material quantities involved in both storage and internal or external transfers;
- Operating records that document nuclear material movements, locations, item identities, item histories, and links to the measurement data; and

- Accounting control procedures that include frequent cross-checking of data to detect inconsistencies, which may be indications of human error, unidentified measurement errors, or an attempt to divert or steal material.

The accounting system may divide a facility or process into sections, which are frequently called material balance areas (MBA) for monitoring and tracking of nuclear material (NM). Measurement, storage, and transfer records are maintained for each MBA. The choice of MBA boundaries will impact detection probabilities and should be carefully considered. The MBA will include defined key measurement points (KMP), where NM is in a form that can be measured to determine material flow or inventory. KMPs include, but are not limited to, all of the inputs and outputs to the MBA. Selection of appropriate KMPs is a fundamental requirement for timely detection of diversion or theft attempts. Ideally these KMPs would be identified during facility or process design to optimize their effectiveness while minimizing the impact on operational efficiency.

One of the primary goals of an accounting system is to monitor the inventory difference (ID) or material unaccounted for (MUF) within a facility or a specific area within a facility, such as an MBA, over a given period of time. The period of time used is typically the time between physical inventories of SNM within the facility, usually one or two months. The ID for SNM over this time period of interest, represented by t , is defined as:

$$ID_t = BI_t + R_t - EI_t - S_t \quad (3.1)$$

where BI is the beginning inventory of SNM, R is receipts of SNM during the inventory period, EI is the ending physical inventory of SNM, and S is shipments out of SNM during the inventory period. Note that all terms may have multiple individual components and include some uncertainty due to measurement and sampling errors. IDs are charted as a function of time and monitored for trends that may indicate a loss of nuclear material. If the safeguards practitioner understands the uncertainties in the components of each of the inputs to equation 3.1, the practitioner may propagate these uncertainties to determine the variance of the ID, σ_{ID} . The monthly ID may then be compared to

statistical limits to determine whether or not an indicator of theft or diversion may exist. An ID in excess of a significant quantity (SQ) indicates an abrupt loss of NM. Trends including systematic, smaller losses of NM over a period of time indicate a protracted loss of material. In some cases, protracted losses may be referred to as a trickle diversion.

A common performance measure for an accounting program is the magnitude of the ID variance, σ_{ID} . IDs must be maintained at a level sufficiently small to detect abrupt diversions, while a sequential test or alternative statistical test for trend detection may be used to detect both abrupt and protracted diversions [3–6]. Specific statistical tests are well documented in the literature and beyond the scope of this section. Safeguards practitioners must understand several issues that complicate the evaluation of IDs, including (1) the impact of holdup on σ_{ID} ; (2) serial correlation in successive IDs; (3) poorly characterized measurement quality and its impact on estimates of σ_{ID} ; and (4) time delays in the measurement results that are input components to the terms in equation 3.1. Without a detailed understanding of these effects, the practitioner will not be sensitive to the subtle variations and trends, the impact of nuclear process changes, or the impact of safeguards policy changes.

Nuclear Material Measurement Technology

The accurate quantification of nuclear material is essential to nuclear safeguards and as noted previously, the minimization of precision and bias in these measurements of NM is vital to the system's ability to detect abrupt and protracted diversion attempts. Measurement of NM is typically accomplished by either destructive wet chemistry techniques or nondestructive measurement of radiation or heat output. Nuclear material is found in multiple chemical forms, a wide range of quantities, and a variety of packaging configurations. Each of these variables introduces uncertainty into measurement results that must be carefully managed and considered when determining whether or not nuclear material is present and accounted for appropriately without evidence of diversion, theft, or misuse.

Destructive Analysis Analytical chemistry typically provides the most accurate techniques for the

quantification of pure metals and compounds of safeguards interest. These techniques require the destruction of a small sample of the nuclear material item, which must be homogenous to prevent the introduction of significant uncertainties from a sample that is not representative of the bulk item. The nature of these destructive analysis (DA) techniques requires ready access to the material to be analyzed. Finished material, such as a fuel rod, is not a candidate for DA for this reason. Some of the more common techniques, which are described below, are gravimetry, titrimetry, and mass spectrometry [7, 8].

Mass spectrometry is a mature technology used for the determination of the isotopic composition of nuclear material samples. Samples to be analyzed in a thermal ionization mass spectrometer (TIMS) are deposited on a filament that is then inserted into the spectrometer and slowly heated by an electrical current. This results in ions of the sample “boiling off” after which they are accelerated by an electric field through a magnetic field, which is orthogonal to their trajectory. This results in a deflection of the ions, the magnitude of which is a function of the weight of the ion. Following the deflection, ions are collected at positions of interest that allow the analyst to infer the isotopic mass of each ion. The precision of this technique for the analysis of plutonium or uranium isotopics is better than 0.05% [7, 8]. Gas mass spectrometers and Isotope Dilution Mass Spectrometers (IDMS) are also in use and in many cases may lead to even better precision than TIMS.

Plutonium or Uranium can be burned to PuO_2 and U_3O_8 , respectively, and then accurately weighed. This technique, called gravimetry, can be performed in the laboratory with random and systematic errors of less than 0.05% [7, 8]. Nonvolatile impurities must be accurately determined and the gravimetric results then corrected accordingly.

Titration procedures are used to assay the amount of uranium or plutonium in solution through the careful addition of a reagent that reacts with the nuclear material. The measured amount of reagent to produce a particular endpoint is then related to the amount of NM. A typical endpoint is detected by a color change or electrical response of the solution. One of the most common techniques in use by the IAEA and other laboratories is the Davies–Gray method [7]. The precision for these techniques can be as good as 0.02% for uranium and 0.04 % for plutonium [7, 8].

DA techniques in use by the IAEA’s Safeguards Analytical Laboratory (SAL) and its network of approved analytical laboratories are summarized in Table 2. In addition to requiring access to the material and destruction of a sample, DA techniques, in general, are more costly and time consuming than nondestructive techniques.

Nondestructive Assay Many nuclear materials that one may encounter, such as a fuel assembly or a radioisotopic heat source, are technologically complex and have high economic value. Nevertheless, to

Nuclear Safeguards and Proliferation of Nuclear Weapons Materials. Table 2 IAEA destructive analysis techniques [7]

Technique	NM	Material Type	Random Error (%)	Systematic Error (%)
Davies–Gray	U	U, MOX	0.05	0.05
MacDonald–Savage	Pu	Pu materials	0.1	0.1
Controlled Potential Coulometry	Pu	Pure Pu	0.1	0.1
Ignition Gravimetry	U, Pu	Oxides	0.05	0.05
K X-ray Fluorescence	Pu	Pu materials	0.2	0.2
Isotope Dilution Mass Spectrometry (IDMS)	U, Pu	Pu, MOX, spent fuel	0.1	0.1
Pu(IV) Spectrophotometry	Pu	Pu, MOX	0.2	0.2
Alpha Spectrophotometry	Np, Am, Cm	Spent fuel	5.0	5.0
Thermal Ionization Mass Spectrometry (TIMS)	U, Pu	Pure U, Pu	0.05	0.05

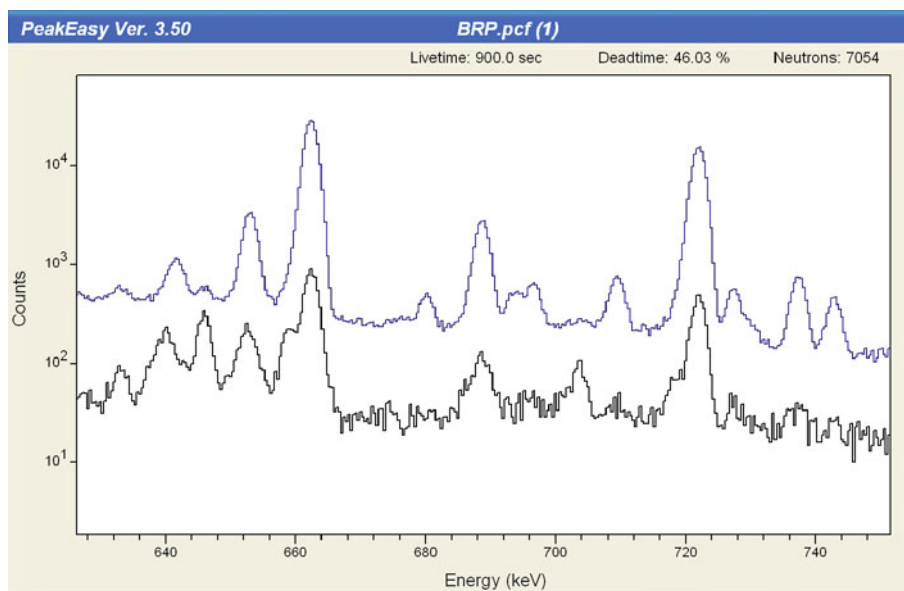
adequately track and account for these materials, they must be quantified accurately. This necessitates the use of methods that do not alter their physical form or integrity of the item. Nondestructive assay (NDA) techniques can provide the necessary information and, although the precision and accuracy is generally poorer than DA, NDA is usually less expensive and can be performed within quicker time periods. DA and NDA should be viewed as complimentary techniques that are both necessary for a quality safeguards measurement program.

Most NDA techniques measure spontaneous or stimulated radiation or heat generated as a result of the radioactive decay process. Passive NDA methods measure radiation that is spontaneously emitted, while active techniques measure radiation that is stimulated by external neutron or gamma ray irradiation. Of the principal radiations, alpha and beta particles do not penetrate bulk material sufficiently to be of use to the NDA professional. On the other hand, x-rays, gamma rays, and neutrons are all electrically neutral, penetrate bulk materials, and are regularly used for NDA applied to nuclear safeguards.

Gamma rays and x-rays are photons with energies above that of the visible light spectrum. X-rays are emitted as result of changes in the electronic structure

of the atom, while gamma rays are emitted when there is a change in the state of the nucleus. In commonly assayed nuclear materials, x-rays in the energy range 80–120 keV or gamma rays from 60–2,600 keV may be used for quantitative measurements. Since every isotope emits radiations at very precise energies and at well-characterized intensities, it provides a signature that is unique to that specific isotope and related to the amount of material present.

One of the most common NDA measurements is to determine the isotopic composition of plutonium or uranium by measuring the intensity of specific gamma rays produced by one isotope of interest relative to those from other isotopes present in the sample. For example, Fig. 1 shows spectra from two types of plutonium. If the spectra are carefully examined, one can see that the characteristics of each spectrum are highly dependent on the isotopic composition of the plutonium. Common instruments for the measurement of gamma ray spectra are scintillators, such as NaI, or semiconductors, such as high-purity germanium. Scintillators are materials that produce optical light pulses when the detector material interacts with a gamma-ray photon. The production of light is proportional to energy deposited in the detector material by the interacting gamma ray. Alternatively, when a gamma



Nuclear Safeguards and Proliferation of Nuclear Weapons Materials. Figure 1

Simulated High Purity Germanium spectra for weapons-grade and reactor-grade plutonium [9]

Nuclear Safeguards and Proliferation of Nuclear Weapons Materials. Table 3 Commonly used gamma detector materials [10]

Material	Type	Average Z	Density (g/cm ³)	Resolution % FWHM @ 662 keV	Resolution % FWHM @ 122 keV
Bi ₄ Ge ₃ O ₁₂ (BGO)	Scintillator	28	7.1	12	28
NaI:Tl	Scintillator	32	3.7	7	13
LaCl ₃ :Ce	Scintillator	28	3.9	3.3	8
LaBr ₃ :Ce	Scintillator	41	5.3	2.8	7
Xe	Gas	54	0.4–0.5	2	7
CdZnTe	Solid-state	49	6.0	3.2	6.3
CdTe	Solid-state	49	6.1	0.6	1.5
Ge (thick planar)	Solid-state	32	5.3	0.2	0.4

photon interacts with a semiconductor (solid-state) material, it creates electron–hole pairs in proportion to the deposited energy. The characteristics of commonly used detector materials are delineated in Table 3. Resolution describes how well a given spectrometer can distinguish one gamma ray from another as a function of energy.

Under some conditions, the measurement of gamma rays may be used to determine the mass of NM present. Most nuclear materials have high atomic numbers and densities and therefore readily scatter and absorb their own radioactive emissions. By measuring the transmission of known gamma-ray intensities at one or more energies, the NDA professional may calculate correction factors to account for the scatter and absorption in the item. It is common practice to use the 413.7 keV gamma ray from plutonium or the 185.7 keV gamma ray from uranium for mass measurements. It should be noted that the self-attenuation factor for plutonium is smaller than that for uranium, since the plutonium gamma ray of interest has a higher energy. This means that self-attenuation corrections for mass measurements of plutonium items are generally smaller than those required for uranium items.

Since many items to be assayed may have the nuclear material dispersed nonuniformly in a matrix that may also be nonuniform, systems that make attenuation correction frequently have to divide the measured item up into a number of elements that are measured individually and have attenuation correction

factors determined for each individual element. Early systems used a technique called Segmented Gamma Scanning (SGS) [11]. The SGS measured the item as a series of horizontal slices; each slice was characterized individually and then summed to arrive at an overall result. With the powerful computers that became economically available in the early 1990s, a tomographic gamma scanner (TGS) was developed that divides an item up into a large number of volume elements [12]. The techniques used to calculate the absorption-corrected mass of nuclear material following the TGS measurement are similar to those used in computed tomography for medical applications. The accuracy of assay results for low-density matrices, such as combustible waste, is as good as 2–3%, but may be as high as 15–18% for higher density matrices such as sludge.

Like gamma rays, neutrons are electrically neutral, but they penetrate high-Z, high-density materials further than a gamma ray of the same energy. Therefore, the measurement of neutrons for the assay of special nuclear material is advantageous for large items and items in high-density matrices. At this point in time, neutron detectors in use around the world do little to distinguish one neutron from another of a different energy. However, neutron intensity and the time correlation of multiple neutron detections provide information that can be used to determine the amount of nuclear material present in a sample if the isotopic composition of the sample is already known. Most nuclear material assay applications in use today using

neutrons, measure coincident neutrons that result from induced or spontaneous fission.

Most neutron detectors use ^3He gas proportional counters embedded in polyethylene. The ^3He has a large cross section (probability) for capturing thermal neutrons. The polyethylene acts as a moderator to slow down the neutrons, produced by fission or (α, n) reactions, to thermal energies where they have a higher probability of interaction with the ^3He gas. The capture interaction then results in the release of a triton and proton that share 765 keV of energy, which then ionize the gas and create an associated electronic pulse that can be measured. Passive detector systems developed using this basic technique have been installed in nuclear facilities for nuclear safeguards of plutonium that have demonstrated precisions of better than 2% [8, 13]. Plutonium measurement is the primary application of passive neutron counting since plutonium has a high rate of spontaneous fission from the decay of its even-numbered isotopes. Counters using higher-order coincident neutrons from fission have also been developed and are in widespread use [14, 15]. Such multiplicity counters are advantageous for the measurement of impure plutonium-bearing items [16]. In general, the precision and accuracy of the measurements using multiplicity counting can be as good as 1–3% [8].

Uranium, due to its long radioactive decay half-lives, emits very few neutrons from spontaneous fission or (α, n) reactions, unlike plutonium. Therefore, effective assay of uranium via neutron measurement requires that the fission rate in the item be externally enhanced. By placing an external neutron source near the item, fissions will be induced in the sample. If the external source used to interrogate the sample produces uncorrelated neutrons, such as those from an AmLi source, they can be distinguished from those produced by the induced fissions in the sample, since these later neutrons will be correlated in time. Alternatively, one can use an external fission source, such as a small ^{252}Cf source, that can be moved away from the sample after inducing fissions or be turned off, such as a small accelerator. In these later two cases, the measurement of neutrons from the induced fissions is then conducted after the external source is no longer present.

The Active Well Coincidence Counter (AWCC), which is in common use around the world, uses AmLi

sources to induce fissions in a uranium sample [17]. Its precision and accuracy have been reported to be better than 5% for the measurement of uranium [8]. Another advantage of using AmLi sources is that it produces neutrons with a mean energy of 0.5 MeV, which is below the fission threshold of ^{238}U . This allows the effective assay of ^{235}U in a sample of uncertain enrichment.

The Californium Shuffler is an example of an instrument that rapidly moves a spontaneous fission source away from a sample to allow subsequent measurement of the induced fissions [18, 19]. This instrument measures the delayed neutrons from fission. Approximately 1.6% of neutrons from fission reactions are produced a few seconds to minutes after the fission. For materials that are well represented with appropriate calibration standards, the shuffler technique has been reported to accuracy as good as 1–2%.

The differential die-away technique (DDT) was developed to use a pulsed accelerator producing 14 MeV neutrons as the external source. This technique has been successfully applied to the assay of both uranium- and plutonium-bearing wastes [20–22]. The combined thermal epithermal neutron (CTEN) method is similar to the DDT method, but interrogates the sample with both thermal and epithermal neutrons [23]. This is achieved partly by the addition of ^4He detectors, which have a faster response than the ^3He detectors and can detect fast fission neutrons in the presence of the epithermal interrogating flux. ^4He detectors rely on the recoil of light nuclei to ionize the gas in the detector tube, which is dependent on the elastic scattering cross section. Because epithermal neutrons are more penetrating in fissile material than thermal neutrons, the differential response can be analyzed to detect the occurrence of self-shielding by fissile material and measure the size of the effect. Self-shielding occurs when discrete lumps of fissile material are present, and can result in assay errors of several hundred percent. The DDT and CTEN methods have been applied to the measurement of SNM in waste containers and for the detection of the potentially smuggled SNM in packages [24].

Many nuclear materials, including plutonium and to some extent uranium, emit heat produced as a result of alpha particle absorption in the sample following radioactive decay. Plutonium produces 2–12 W of heat

for each kilogram of material in the sample depending on the specific isotopic composition of the plutonium. Radiometric calorimeters measure the heat output of nuclear materials [25]. This method has seen wide application to plutonium and tritium samples, as well as large (kilogram) quantities of uranium and neptunium [26, 27]. Calorimetry is generally considered the most accurate and precise NDA technique for plutonium measurement since it is not subject to many of the biases that effect both gamma-ray and neutron measurements. Accuracy and precisions better than 1% are generally reported in the literature [28]. As is required for neutron-based NDA, the calorimetry data must be combined with isotopic data from gamma-ray or mass spectroscopy to obtain the total amount of plutonium in a specific item. Although calorimeters are the most accurate technique, it usually requires long measurement times (typically 4–8 h) and is not generally portable. These disadvantages make the application of calorimeters to international safeguards and inspection efforts very limited.

Detection of Nuclear Proliferation and Treaty Verification

There is a well-established historical record of states pursuing nuclear weapon development, even after accepting international obligations through treaties not to develop weapons. In recent years, those countries have included Iraq, Libya, and North Korea. It is likely that such trends will continue in the future. This necessitates that the international community have the capability to detect, deter, and even dismantle undeclared nuclear capabilities. It should be noted that Libya agreed to abide by the NPT in 2003 and following negotiations with the USA and Great Britain, the USA airlifted out of Libya components of the abandoned nuclear weapons program in 2004.

Open Source Information and Satellite Imagery

The analysis of open source information and commercial satellite imagery are two methods currently in use by the IAEA and state intelligence assets. Open sources of information include such things as publically available data, media reports, published scientific literature, business reports, information on technical conferences and meetings, and patent applications. Open

sources of information are numerous and continuously changing. Each type of information must be assessed in relationship to the others in an attempt to establish an integrated picture of a state's legal nuclear programs, undeclared activity, and potential to breakout and quickly establish a nuclear weapons capability. The open source analyst, which should have a broad technical background, must understand the potential quality pitfalls from the large volume of data available and then formulate useful search, query, and analysis strategies. This type of analysis requires sustained efforts over time, personnel with excellent research skills, and teams with multidisciplinary backgrounds [29].

Another open source is commercial satellite imagery, which has recently become available for anyone with modest computer access to view, identify, and monitor facilities that may be used in a nuclear weapons program. Commercial satellite images have been available since the 1970s, although at that time resolution was as poor as 80 m. Today, commercial images with resolution better than 0.5 m are becoming widely available for panchromatic data [30, 31]. Commercial satellites typically acquire images in the visible and/or near-infrared parts of the electromagnetic spectrum. Multispectral satellites, which collect data from more than one electromagnetic band, generally have poorer resolution (~ 2.5 m) than those collecting panchromatic data (black and white). An image that results from the fusion of the multispectral data and a panchromatic overlay is generally the best for facility and infrastructure analysis.

Imagery analysis is the process of identifying, analyzing, and attaching significance to the image data. An analyst's effectiveness requires

- Size – an ability to assess true and relative size of the objects being imaged
- Shape – an ability to determine physical characteristics
- Shadows – an ability to use the sun angle and orientation on the imaged objects
- Shade – an ability to use tonal brightness and contrast of objects of interest
- Surroundings – an ability to use context and setting of the objects including topographic and geographic information

- Signatures – an ability to use cultural or manmade features to help separate objects from their surroundings and to uniquely identify them
- Time – an ability to monitor temporal changes to provide insight on functionality, operating histories, or facility changes [32]

These techniques have been used multiple times in the detection and monitoring of clandestine nuclear activities. They are complementary to direct access inspections, open source analysis, and traditional nuclear safeguards (nuclear material accounting and measurements) in the detection and deterrence of nuclear weapons development.

Seismic and Hydroacoustic Monitoring

As a nation tries to develop new nuclear weapons, whether or not it already possesses one or more earlier generation nuclear weapons, they are likely to test new designs. It is probable that such tests would be conducted underground to both contain the fission products and conceal the exact nature of the test from other countries. Hence, seismic monitoring is an important tool in the nonproliferation regime [33]. Monitoring requires the detection of the signal, separation of the signal from background and natural events (earthquake vs. explosion), as well as estimating the location and size of the event. There are more than 200,000 seismic events that are similar in magnitude to a small nuclear explosion every year. These events are either earthquakes or chemical explosions associated with peaceful applications, such as mining.

The energy released by a sudden movement of the earth's crust produces two types of elastic body-waves. The first to arrive at a monitoring station are the high velocity P-waves, which move in a direction parallel to the direction of the wave energy. S-waves move as transverse waves with motion perpendicular to the direction of wave propagation and arrive at the monitoring station at a later time relative to the P-wave. An explosion produces little S-wave energy relative to that produced by an earthquake. Therefore, the shape of the wave recorded by a monitoring station can be used to distinguish between an explosion and earthquake. The Comprehensive Test Ban Treaty, which was opened for signature in 1996, specifies a minimum detection level for explosions of 1 kiloton. This level,

which is currently exceeded by international monitoring capabilities, corresponds to a seismic event of approximately 4.0 on the Richter scale. At some locations around the globe, monitoring capability is better than 2.5 on the Richter scale. How this relates to the size of a nuclear explosion is dependent on the geology between the event and the monitoring stations [34, 35].

Since more than 65% of the earth's surface is water, it is also important to monitor for explosive events in the world's oceans. Underwater explosions produce extremely powerful broadband acoustic signals. A sensitive hydrophone either suspended from a moored buoy or anchored to the ocean floor can be used to monitor low frequencies (1–100 Hz). Such low-frequency monitoring is optimal for detecting explosions over long distances. Today's hydrophones can detect signals more than 15,000 km from their source. The waveform detected can be used to differentiate explosions from underwater earthquakes or volcanoes.

Environmental Surveillance and Monitoring

Another tool that can be used to monitor treaty compliance or breakout activity by a proliferant is the surveillance of radionuclides present in the environment [36]. The fission products produced by nuclear fission and their isotopic ratios are an unambiguous and well-characterized indicator. However, unlike seismic or acoustic monitoring, it is not done in real time. Atmospheric and underwater explosions release particles carrying radionuclides as well as noble gases, while underground explosions release noble gases. In either case, the nuclides are then transported by air or water to monitoring stations. At that point in time, which may be several days or even weeks following the event, the analysis of the sample can confirm the nuclear event but not pinpoint the location. Combining seismic and acoustic monitoring with the atmospheric surveillance, with meteorological data, can provide definitive evidence of a nuclear explosion and its location.

Most of the energy of a nuclear explosion is transformed into the immediate blast or the radiation released during the blast. Approximately 10% of the energy is released as residual radiation, which is emitted over time, through the decay of fission products generated during the nuclear blast. Fission products are produced in both gaseous and particulate form and

most are radioactive. Following an atmospheric explosion, solid fission products attach to dust particles, which may then be propagated by winds over significant distances. Underwater explosions also release particles into the atmosphere. Underground explosions may be conducted in such a way that particles carrying radionuclides are contained. However, some fission products, particularly xenon, are noble gases that are much more difficult to contain and can be collected and analyzed by monitoring stations thousands of kilometers from the explosion site [37].

A radionuclide particulate monitoring station contains an air sampler and detection equipment. At the air sampler, air is forced through a filter, which typically retains 85% or more of all particles that reach it. The used filter is usually first cooled for a period of 24 h and then measured for another 24 h in the detection device at the monitoring station. The result is a gamma ray spectrum that the analyst can use to identify specific radionuclides. The analyst can use the spectrum to determine characteristics of the device tested and the nuclear material used in the device.

Detection and Prevention of Illicit Trafficking and Terrorism

Since the attacks in the USA on September 11, 2001, terrorism's threat against a country's national security has come under increased scrutiny. Although few terrorist organizations will implement acts of such devastation, it is considered by many experts that nuclear terrorism cannot be ruled out, and, because the devastation of a nuclear attack would be unmatched, the risks must be mitigated regardless of the low probability [38–40]. To mitigate these risks, one may increase security at the storage locations (strengthen nuclear safeguards), dispose of the nuclear materials in such a way that they are unavailable to the terrorists, strengthen methods to detect their misuse and transport once they have fallen into the wrong hands, and also develop methods through which illegal activity may be attributed to a particular national or subnational group.

Rigorous nuclear safeguards, which have been described previously, are only one layer in a defensive system against the use of nuclear materials by a subnational or terrorist organization. In addition to

nuclear safeguards, the other layers of mature protection system would include monitoring of nuclear material in transit, robust intelligence collection and analysis, consequence management, and a vigorous response capability to a nuclear emergency or event to mitigate it prior to a detonation. Each of these layers relies on, at least in part, the ability to detect and characterize nuclear or radiological materials at various points in their life cycle.

Nuclear smuggling is difficult to stop due to the small size of the materials required by the terrorist to make a nuclear bomb; the weak radiation many of these materials emit; the large number of smuggling routes; the large amount of legitimate traffic crossing these borders every year; and the existence of established smuggling networks for other contraband, such as illegal drugs. A recent analysis of 174 illicit radiological material trafficking incidents that have occurred since 1991 found little evidence of clear trends that can be used by nuclear security personnel [41]. It did note that the incidents involved a variety of radionuclides; however, approximately 60% of the cases involved Cesium-137, which happens to be an attractive isotope for radiological dispersal devices.

Border and Port Monitoring

Studies of illicit trafficking of nuclear and radiological materials have indicated that whether by auto, bus, train, or air, smugglers have consistently used transportation methods and routes that were most convenient to them, maintaining a close physical proximity to the materials [42]. During the past decade, one of the areas that has received international emphasis for this very reason is the improvement of border and port detection capability and protocols.

In the USA, one such effort is the Second Line of Defense program at the Department of Energy's National Nuclear Security Administration. This program seeks to interdict illicit trafficking of nuclear and radiological material through airports, seaports, and border crossings in Russia and other key transit states. This is to be accomplished by helping other nations install and use radiation detection equipment at these sites and by providing associated training and support. It is worth investing in improved border detection systems, but this line of defense will inevitably be porous, and one should not place undue reliance on it.

Fixed portal monitors have been installed at land, sea, and air facilities with the goal of detecting smuggled radioactive materials. Typically, these detectors use passive gamma measurement, which in its simplest and least expensive configuration utilizes gross gamma counting with little discrimination of the gamma-ray energy. Plastic scintillators are the most common detector material for this application and are inexpensively available in large sizes. Pedestrians, vehicles, or cargo pass through one or more pairs of these detectors. The number of gamma rays detected over the time interval during which the object is between the detectors can be used to trigger an alarm requiring the intervention of a guard or similar security personnel. There is a significant trade-off between the sensitivity of this method and the permissible duration that the object can be measured. Also, if the sensitivity is set such that frequent false alarms occur, security personnel may ignore alarms in order to keep traffic moving efficiently. Most of the radiation emitted by highly enriched uranium (HEU) is weak, which means it is both easy to shield by the smuggler and difficult to detect through passive means. The detectors being deployed in the Second Line of Defense program simply detect gamma-ray radiation, without attempting to zero in on particular energy levels, and therefore have almost no ability to detect HEU above the background of natural radiation if the terrorist or smuggler uses even modest radiation shielding.

Efforts are underway to develop “advanced spectroscopic portal monitors” (ASPs) that do examine the energy ranges of different gamma rays. This will theoretically allow them to not only detect nuclear or radiological material, but also to identify the type of material. The ASPs are expected to reduce the rates of false alarms caused by naturally radioactive materials in legitimate shipment; however, they are significantly more costly than nonspectroscopic detectors and can only be manufactured with much lower detection probabilities. Thus, substantial controversy exists over whether there is a true benefit to their deployment.

Active detectors, such as neutron sources coupled with detectors to detect induced fission, may offer higher confidence of detecting HEU, even in a shielded configuration. Several methods are under active research and development that have been demonstrated to be highly effective at detecting shielded

SNM. These detection methods are based on the active neutron techniques developed for nuclear safeguards discussed previously. Other development efforts are underway to use high-energy photons to induce fission. In both cases, neutron or photon sources, systems are in the developmental stages and not yet ready for wide-spread deployments.

Nuclear Forensics

In response to nuclear smuggling, which was first reported in 1991, a new branch of science was developed: “nuclear forensics” [43]. Analytical methods were initially borrowed from nuclear safeguards and supplemented with techniques developed for material science investigations. Over several ensuing years, techniques were standardized, new methods developed, and extensive databases of nuclear information compiled to guide the analysis and attribution of nuclear materials. As in other forensic fields, the analyst relies on the fact that certain measureable parameters in a sample are characteristic of the nuclear material and can be used to draw conclusions on the intended use or origin of the materials. Such attribution is an important deterrent to a nation state’s participation, whether willingly or not, in smuggling activity.

The first category of characteristics to be exploited for nuclear forensics was major elemental and macroscopic composition of the interdicted nuclear material. Physical dimensions and isotopic composition of the major nuclear materials, usually uranium and plutonium, may be used to identify the intended use of the material. For example, plutonium with less than 7% ^{240}Pu is considered to be weapons grade and likely produced for that purpose. The isotopic composition may also indicate the origin of the material. Since different types of nuclear reactors have different fuel loadings and neutron energy spectra, they generally produce plutonium of differing isotopic composition [44]. The concentration of the nuclear material may indicate stoichiometry of the compound or the presence of impurities, both of which may be indicators of the materials’ source or status in a known nuclear fuel cycle. Fuel additives, which are used in the fuel cycle or material processing steps, such as gadolinium as a nuclear reactor burnable poison, may also be used by the analyst to narrow down

the type of nuclear material present. In the case of manufactured items, such as nuclear fuel elements or pellets, the physical dimensions may allow the identification of the intended reactor type or design.

Further developments in the nuclear forensics field lead to the use of minor elemental as well as microscopic composition of the nuclear material in investigations initiated in the late 1990s [42]. This included chemical and isotopic analysis of individual particles as part of the forensic toolbox. Techniques such as secondary ionization mass spectrometry and scanning electron microscopy coupled with electron dispersive x-ray analysis allow for the isotopic composition of particles or powders to be determined. With powder samples, the analyst also gains information related to the sample homogeneity with regard to the isotopic composition and the particle size distribution.

In the past decade, analysts have also developed geolocation techniques and methods. Trace elements can be found in uranium at all stages of its mining, milling, and enrichment. Each processing step reduces the concentration of these trace elements and may introduce new trace contaminants. However, it has been noted that the pattern of certain impurities does not vary through the processing and may be used as a characteristic of the starting material. For example, the distribution of lead isotopes in the sample may be used as an indication of the age of the ore body and its initial U/Th ratio. The ratio of $^{18}\text{O}/^{16}\text{O}$ in a uranium oxide sample may be used to determine the temperature, latitude, and distance to the sea for the water used in the wet processing of the uranium [42].

The challenge in the interpretation of the many data sources discussed above is the accessibility of reference data. Efforts are underway to develop databases of reference forensic information. Data on nuclear materials produced for commercial applications, such as power production, are generally available with the various manufacturers and to some extent in open literature. However, some of the data is commercially sensitive and its availability is limited. Also, similar data on nuclear materials developed for weapons applications is generally protected for national security reasons.

Terrorism Response and Consequence Management

If a terrorist succeeds in smuggling nuclear or radiological material through the layers of nuclear safeguards and security at the source and subsequent national or international monitoring (such as the Second Line of Defense) it may then be readied for use as a weapon. If monitoring stations, intelligence, or criminal investigations lead to the discovery of the material or weapon, a response is required to mitigate the threat. That response requires personnel familiar with weapons design, device characterization, weapon defusing and disposal, and consequence management.

To assist the response or management of the consequences of the event, small portable instruments are required for identification and quantification of the materials involved. Many responders, including those that may be on the incident scene first, such as police officers, today use hand-held or backpack size instruments called radioisotope identification devices (RIIDs). These instruments have undergone significant advancements over the past several years in both hardware and algorithms.

The algorithm used for isotope identification by a RIID plays a key role in obtaining a correct identification of the radioisotope. Correct ID is perhaps the most essential step in determining whether or not a true threat exists that requires a more extensive response for mitigation. The majority of RIIDs deployed around the world are based on NaI spectrometers. Their performance in isotope identification is generally poor and therefore secondary analysis of spectra by a trained spectroscopist is frequently necessary to resolve problems in the isotope identification. A trained spectroscopist is capable of identifying complicated, multiple-line sources with even the poorest resolution detectors, such as NaI. Many factors in detector performance, such as energy resolution, play an important role in isotope identification and are active areas of research [45–47].

Future Directions

World energy demand continues to grow and along with it does the demand for nuclear power, a carbon-free energy source. Some are predicting that world energy demand growth will exceed 50% over the next 20–25 years [48]. This expansion presents both

challenges and opportunities to nuclear safeguards and nonproliferation. Challenges include the timely detection of diversion at large throughput fuel cycle facilities, new types of fuel cycles that may present challenges to accurate measurement of the fuel at appropriate KMPs, and the detection of nuclear material misuse in the background of activity required for a large-scale nuclear fuel cycle. Finally, significant expansion of nuclear power will also include increased nuclear material shipping, which creates its own issues related to continuity of knowledge and monitoring of the material in transit.

Addressing these future challenges will require advances in instrumentation, safeguards system analysis, modeling techniques, and data mining. On-line, near real-time monitoring methods using both radiation and nonradiation signatures will be required as part of integrated process monitoring schemes for new facilities. There will also be opportunities to incorporate nuclear safeguards and physical protection into the design of new facilities. This will allow the efficacy of the safeguards system to be maximized while minimizing cost and impact to the facility.

Improvements in the detection of shielded nuclear material is one of the most active areas of research to improve border monitoring and other efforts related to illicit nuclear trafficking. As worldwide commerce continues to increase, so will the movement of goods around the world. This will necessitate that the false alarm rates of the current generation of portal monitors be significantly improved to avoid considerable negative impacts to commerce. Although this is another area of active research, implementable solutions are likely still several years away.

Bibliography

Primary Literature

- Fischer D (1997) History of the International Atomic Energy Agency: the first 40 Years. IAEA, Austria
- Kennedy RF (1968) Thirteen days: a memoir of the Cuban missile crisis. McCall, New York
- Jaech J (1973) statistical methods in nuclear materials control, U.S. Atomic Energy Commission
- Burr T, Coulter C, Hakkila E, Ai H, Kadokura I, and Fujimaki K (1995) Statistical methods for detecting loss of materials using near-real time accounting data. 36th Annual Meeting of the Institute of Nuclear Materials Management, Palm Desert, California
- Picard R (1987) Sequential analysis of materials balances. *J Nuclear Mater Manage* 15(2):38–42
- Zrdecki A, Armstrong J, Longmire V, and Strittmatter R (1997) Inventory difference analysis at Los Alamos Plutonium Facility. 38th Annual Meeting of the Institute of Nuclear Materials Management, Phoenix, Arizona
- IAEA (2003) Safeguards Techniques and Equipment – 2003 Edition. IAEA/NVS/1 (revised), International Nuclear Verification Series No. 1, Austria
- Aigner H, Binner R, Kuhn E, Blohm-Hieber U, Mayer K, Guardini S, Pietri C, Rappinger B, Mitterrand B, Reed J, Mafrà-Guidicini O, and Deron S (2002) International target values 2000 for Measurement uncertainties in safeguarding nuclear materials. *ESARDA Bulletin*, No. 31
- Garner S, (2009) personal communication, July 28, 2009
- Russo P and Vo D (2005) Gamma-Ray Detectors for Nondestructive Analysis, Los Alamos National Laboratory, LA-UR-05-3813
- Martin R, Jones D, Speir L, and Walker A (1974) Field assay of plutonium with a new computerized segmented gamma scan instrument, Los Alamos Scientific Laboratory, LA-UR-74-924
- Estep R, Prettyman T, Sheppard G (1993) Tomographic gamma scanning to assay heterogeneous waste. *Nuc Sci Eng* 118:145–152
- Menlove H, Swansen J (1985) A high-performance neutron time correlation counter. *Nuclear Technol* 71:497–505
- Krick M, Swansen J (1984) Neutron multiplicity and multiplication measurements. *Nucl Inst Methods* 219:384–393
- Langer D, Krick M, Ensslin N, Bosler G, and Dytlewski N (1991) Neutron multiplicity counter development, Proceedings of the 13th Annual ESARDA Symposium, Avignon, France
- Stewart J, Krick M, Langer D, and Wenz T (1998) Neutron multiplicity assay of impure materials using four different neutron counters, 39th Annual Meeting of the Institute for Nuclear Materials Management, Naples, Florida
- Menlove H, Foley J, and Bosler G (1980) Application of the active well coincidence counter to the measurement of uranium, 2nd Symposium on Safeguards and Nuclear Material Management, Edinburgh, Scotland
- MacMurdo K, Bowman W (1977) Assay of fissile materials by a cyclic method of neutron activation and delayed neutron counting. *Nucl Inst Methods* 141:299–306
- Menlove H, Crane T (1978) A ^{252}Cf Nondestructive assay system for fissile material. *Nucl Inst Methods* 152:549–557
- Kunz W, Atencio J, and Caldwell J (1980) A 1-nCi/g Sensitivity transuranic waste assay system using pulsed neutron interrogation, 21st Annual Meeting of the Institute for Nuclear Materials Management
- Caldwell J, Hastings R, Herrera G, and Kunz W (1986) The Los Alamos second-generation system for passive and active neutron assays of drum-sized containers, Los Alamos National Laboratory LA-10774-MS
- Rinard P, Coop K, Nicholas N, Menlove H (1994) Comparison of shuffler and differential die-away technique instruments for the assay of fissile materials in 55-gallon waste drums. *J Nuclear Mater Manage* 22(3):4–28

23. Coop K and Hollas C (1996) Epithermal interrogation of fissile waste, Institute of Nuclear Materials Management Annual Meeting, Naples, Florida
24. Rooney B, York R, Close D, and Williams H (1998) Active interrogation package monitor, IEEE Nuclear Science Symposium, Toronto, Canada
25. Bracken D, Biddle R, Carillo L, Hypes P, Rudy C, Schneider C, Smith M (2002) Application guide to safeguards calorimetry, Los Alamos National Laboratory LA-13867-M
26. Walsh T, Hamilton R, Baker E, Hurlbut S, Fazzari D, Delegard C, McRae L, Liebetrau A, Lemaire R, and DeRidder P (1996) Plutonium assay for safeguards purposes: material heterogeneity and application of calorimetry, 37th Annual Meeting of the Institute for Nuclear Materials Management, Naples, Florida
27. Thornton M, Vassallo G, Miller J, Mason J (1995) Design and performance testing of a tritium calorimeter. *Nucl Inst Methods A* 363:598–603
28. Guardini S (2003) Performance Values for Nondestructive Assay (NDA) techniques applied to safeguards: The 2002 evaluation by the ESARDA NDA Working Group, ESARDA Bulletin 31
29. Wallace R, Anzelon G, Essner J (2009) Safeguards information from open sources. *J Nuclear Mater Manage* 37(4):30–40
30. Hitchens T (2006) European eyes in the sky: strategic independence is focus as security comes to the forefront. *Imaging Notes* 21(3):20–24
31. Niemeyer I (2009) Safeguards information from satellite imagery. *J Nuclear Mater Manage* 37(4):41–48
32. Pabian F (2008) Commercial satellite imagery: another tool in the nonproliferation verification and monitoring toolkit, Chapter 12 of nuclear safeguards, security, and nonproliferation. Elsevier, New York
33. Hannon W (1985) Seismic verification of a comprehensive test ban. *Science* 227(4684):251–257
34. Bache T (1982) Estimating the yield of underground nuclear explosions. *Bull Seismol Soc Amer* 72(6B):S131–S168
35. Taylor S, Dowla F (1991) Spectral yield estimation of NTS explosions. *Bull Seismol Soc Amer* 81(4):1292–1308
36. Yutaka M, Tetsuzo O, Hiroshi N, Hideo N (1998) Overview of atmospheric radionuclide monitoring techniques for CTBT International Monitoring System, Proceedings of the Annual Meeting of INMM Japan Chapter, 19:129–130
37. Carrigan C, Heinie R, Hudson G, Nitao J, Zucca J (1996) Trace gas emissions on geological faults as indicators of underground nuclear testing. *Nature* 382:528–531
38. Ferguson C, Potter W, Sands A, Spector L, Wehling F (2005) The four faces of nuclear terrorism. Routledge, New York
39. Jenkins BM (2008) Will terrorists go nuclear? Prometheus Books, New York
40. Allison G (2005) Nuclear terrorism: the ultimate preventable catastrophe. Holt Paperbacks, New York
41. Kerst R (2009) Trends and patterns in illicit trafficking of radioactive sources, LLNL-GS-0052-2009, Lawrence Livermore National Laboratory
42. Battelle Memorial Institute (2009) Nuclear smuggling handbook: case studies and detection methods
43. Mayer K, Wallenius M, Fanghanel T (2007) Nuclear forensic science – from cradle to maturity. *J Alloys Compounds* 444–445:50–56
44. Wallenius M, Lutzenkirchen K, Mayer K, Ray I, Aldave de Iss Heras L, Betti M, Crombroom O, Hild M, Lynch B, Nicholl A, Ottmar H, Rasmussen G, Schubert A, Tamborini G, Thiele H, Wagner W, Walker C, Suleger E (2007) Nuclear forensic investigations with a focus on plutonium. *J Alloys Compounds* 444–445:57–62
45. Sullivan C, Garner S, Lombardi M, Butterfield K, Smith-Nelson M (2007) Evaluation of key detector parameters for isotope identification. *IEEE Nuclear Sci Symposium Conf Record* 2:1181–1184
46. Keyser R, Rwomey T, Upp D (2005) A Comparison of an HPGe-based and NaI-based Radioisotope Identifier (RIID) for radioactive materials. ESARDA Symposium, London
47. Sullivan C, Garner S, Blagoev K, Weiss D (2007) Generation of customized wavelets for the analysis of gamma-ray spectra. *Nuclear Instruments Methods A* 579(1):275–278
48. Energy Information Administration (2007) International Energy Outlook 2007, DOE/EIA-0484

Books and Reviews

- Doyle JE (ed) (2008) Nuclear safeguards, security, and nonproliferation: achieving security with technology and policy. Elsevier, New York
- Hakansson A, Jonter T, (2007) An Introduction to Nuclear Non-Proliferation and Safeguards. SKI Report 2007:44
- IAEA, (2007) Combating illicit trafficking in nuclear and other radioactive material: reference manual, IAEA Nuclear Security Series No. 6. International Atomic Energy Agency, Vienna
- Moody K, Hutcheon I, Grant P (2005) Nuclear forensic analysis. CRC Press, Boca Raton
- Mozley RF (1998) The politics and technology of nuclear proliferation. University of Washington Press, Seattle/London

Nuclear Transfer to Produce Transgenic Mammals

ALEXANDER KIND, ANGELIKA SCHNIEKE
Chair of Livestock Biotechnology, Technical University of Munich, Freising, Germany

Article Outline

Glossary
Definition of the Subject
History and Introduction

Nuclear Transfer
Future Directions
Bibliography

Glossary

Cell cycle The sequence of stages between one cell division and the next. These are: gap 1 (G1) DNA synthesis (S), gap 2 (G2), and mitosis (M).

Chimera An organism composed of a mixture of genetically distinct cells that originate from two individuals.

Chromatin The natural form of genomic DNA within the nucleus. A highly structured multi-coiled fiber composed of a complex of DNA, RNA and proteins.

Clone (a) Molecular clone, an isolated DNA fragment propagated artificially in bacteria. (b) Cell clone, a group of genetically identical cells descended from a single individual. (c) Animal clone, an animal produced by embryo splitting or nuclear transfer.

Cytoplasm The portion of a eukaryotic cell outside the nucleus.

Differentiation A process by which a cell takes on a more specialized role or function.

Epigenetic regulation Regulation of gene expression by modification of chromatin not involving changes to the DNA sequence, e.g., by methylation of cytosine bases.

Locus A defined position in the genome.

Metaphase A stage of cell division when chromosome pairs are condensed and ready to be divided between the daughter cells.

Pluripotent The ability of a single cell to generate all cell types of the body. Pluripotent cells are distinct from totipotent cells in that they cannot make extraembryonic structures, such as the amniotic sac or the placenta.

Promoter trap A method of enriching gene-targeted cell clones against a background of random integrants. Expression of a selectable marker is conditional on integration at the desired location and driven by the promoter of the target gene.

Pronuclei Two structures formed from the sperm and oocyte genetic material following fertilization. Pronuclei later fuse to form the nucleus of the zygote.

Reprogramming A general term describing a radical change in the pattern of gene expression by a nucleus. This may be in response to a change in cytoplasmic factors, for example, after transfer of a nucleus from a somatic cell into an oocyte.

Somatic cell Cells of an organism other than the germ line.

Transcription The first stage in the expression of a gene, in which an RNA copy, or transcript, is made from a DNA template.

Transfection Term covering a variety of chemical, electrical and mechanical methods of introducing nucleic acids into cells.

Undifferentiated Relating to a cell that is capable of differentiation but has not generated specialized characteristics.

Zygote The one-cell embryo formed after fusion of the male and female pronuclei.

Definition of the Subject

Over the last 3 decades, researchers in molecular genetics and developmental biology have generated a repertoire of powerful biological techniques that allow genes to be isolated, analyzed, modified at will, then transferred and studied in cultured cells and live animals. This technology has revealed the functions of thousands of genes and dramatically advanced the knowledge of normal and disease states, providing the basis for numerous practical benefits especially in biomedicine. This entry tracks one strand within this interconnected and ever-broadening field: the use of nuclear transfer to generate transgenic mammals.

Transgenic animals were originally defined as those that contain additional DNA, such as a gene introduced from another species. The term “transgenic” is now commonly used to encompass a range of experimentally engineered genetic modifications including DNA deletion, rearrangement, and replacement. Here the word is used in this broader sense.

Nuclear transfer is the replacement of the nuclear genetic material of an oocyte or zygote with the nucleus of another cell. The technique has been applied in three main areas:

1. Therapeutic cloning – the generation of embryos and stem cells from differentiated cells.

2. Reproductive cloning – the replication of animals from cells, for example, where sexual reproduction is difficult or impossible, or to preserve valuable genetic material.
3. Cell-mediated transgenesis – the production of transgenic animals from genetically modified cultured cells.

The production of transgenic animals was a principal motivation for the development of nuclear transfer in mammals and is the focus of this entry.

This entry is structured as follows: Section “[History and Introduction](#)” gives some historical background to the conception and development of nuclear transfer. Section “[Nuclear Transfer](#)” outlines the current state-of-the-art and related transgenic technologies. Section “[Future Directions](#)” describes possible developments in the near future. This entry can however provide only an overview of this diverse and expanding subject. Interested readers can gain more detailed information by accessing the key papers, reviews, and books cited.

History and Introduction

Cloning animals by nuclear transfer is not a new idea. The first recorded reference is attributed to Yves Delage, a French marine biologist who wrote in 1895: “if, without any deterioration, the egg nucleus could be replaced by the nucleus of an ordinary embryonic cell, we should probably see this egg developing without changes” [1]. There is however no evidence that Delage actually carried out such an experiment. Hans Spemann described the first nuclear transfer in 1928 [2]. In a remarkable piece of microsurgery using micro-tweezers and a loop of hair from his baby daughter, he constricted a single-cell salamander embryo into two parts, one of which contained the cell nucleus. Left to develop, the portion with the nucleus divided and formed an embryo, while the other side remained a pouch of clear cytoplasm. As the embryo developed further, Spemann allowed a single nucleus to pass back into the empty cytoplasm; this reconstructed single cell then developed into another normal embryo.

In the early 1950s, Robert Briggs and Thomas King performed a series of nuclear transfer experiments with frogs. They removed the nucleus from an activated frog oocyte using a glass needle. A single cell dissected from

a later stage embryo was then drawn up into a fine glass pipette connected by rubber tubing to a syringe. The cell broke open as it was squeezed within the pipette, and the free nucleus was then injected into the enucleated egg. Culturing the reconstructed embryos further, they found that cell nuclei from blastula stage embryos could support normal development to tadpoles. But nuclei from later stage embryos, in which the major embryonic cell lineages were already established, were unable to do so.

Transfer of nuclei following this basic scheme was continued by various researchers for several decades. The work was at first restricted to the large eggs of amphibians, but in the late 1970s, embryo culture and micromanipulation techniques improved sufficiently to use the far smaller and more vulnerable eggs of mammals. The overwhelming finding was that the developmental capacity of transplanted nuclei decreased with the age and extent of differentiation of the donor cell. Nuclei of very early embryonic cells had equivalent potential, but at some stage in development, it seemed that the fate of different cells became determined, “hard wired” in some way. The mechanism was unknown, but some form of irreversible modification or loss of nuclear DNA was viewed as a likely explanation. The concept of cell determination thus became widely accepted among developmental biologists. In 1984, James McGrath and Davor Solter seemed to put an end to the possibility of nuclear transfer in mammals. They systematically transferred nuclei from one-, two-, four-, eight-cell and blastocyst stage mouse embryos into enucleated one-cell stage embryos, zygotes. Nuclei from one-cell embryos supported development to blastocysts, but success dropped off sharply using two-cell stage nuclei and failed entirely with later stages. They interpreted this as a rapid loss of nuclear capability during development and concluded their paper with the categorical statement: “the cloning of mammals by simple nuclear transfer is biologically impossible” [3].

However, strict cell determination remained difficult to reconcile with the well-known capacity for organ regeneration shown by various vertebrates, including most fish and some types of amphibian. If for example a newt loses a limb, cells from surrounding tissues migrate into the wound and undergo a process of reverse development, changing their identity to form

a mass of rapidly dividing undifferentiated cells termed a blastema. Cells within the blastema then organize and undergo processes similar to those in embryo development to form a replacement limb [4]. This was good evidence that some adult differentiated cells are not determined in their fate and can radically change their identity. So, despite the repeated failure of experimental nuclear transfer, there remained the tantalizing possibility that the limitations were essentially technical rather than biological. This provided sufficient motivation for some researchers to continue testing.

In retrospect, it was unfortunate that early efforts focused on mice, because they are now known to be one of the more difficult species to clone by nuclear transfer. So, somewhat unusually, the major breakthroughs were made with livestock. The first indication that nuclear transfer from later stage cells might be possible came ironically just a few months before McGrath and Solter's paper was published. Steen Willadsen produced three Suffolk sheep by merging single cells from eight-cell embryos with enucleated unfertilized eggs [5]. It later emerged that the reason for the discrepancy with the mouse work was because McGrath and Solter had used enucleated zygotes for nuclear transfer, because mouse oocytes are too fragile to survive nuclear transfer. Willadsen had been able to use unfertilized oocytes, which are more robust in sheep. Years of work have subsequently shown that unfertilized oocytes are successful recipients for nuclear transfer in numerous species, while zygotes can only be used at a very particular stage.

In the decade that followed Willadsen's achievement, nuclear transfer remained restricted to cells of the very early embryo. While this was potentially useful in increasing the number of embryos of valuable animals such as prize cattle, there were few other practical applications. However, developments in transgenic technologies over the same period led to increasing pressure to improve nuclear transfer.

In 1980, it was reported that naked DNA microinjected into the pronuclei of fertilized mouse eggs could stably integrate into the host genome [6]. Microinjection by the same procedure was extended to livestock 5 years later [7]. A host of practical applications for transgenic livestock species were soon envisaged, including enhanced production characteristics, disease resistance, and the production of

pharmaceutical proteins. However as time went on, it became apparent that DNA microinjection suffered numerous drawbacks. The process was inefficient and offered no control over where a transgene would integrate in the host genome. Randomly integrated transgenes exhibited a very wide range of expression, and this was a major problem for biotechnology companies seeking to make useful transgenic products, such as pharmaceuticals.

A quite different means of generating genetically modified animals, based on the use of cultured cells, was also developed during the 1980s. In 1981, Martin Evans and Gail Martin independently isolated embryonic stem (ES) cells from early mouse embryos [8, 9]. ES cells can be grown indefinitely in culture and then be incorporated back into a developing embryo. In 1984, it was found that these cells could also contribute to the germ line and produce functional sperm or oocytes [10]. This provided a means of establishing mouse strains derived from the ES cell genotype, including any experimental modifications. In 1987, a method of introducing predefined alterations into genes *in situ*, originally developed in tumor cell lines [11], was applied to ES cells by Thomas and Capecchi [12]. This technique, termed gene targeting, made it possible to engineer and study precise genetic alterations in whole animals. The phenomenal power of gene targeting in ES cells has provided an abundance of knowledge about the function of genes in mammals, and was recognized by the 2007 Nobel Prize for Medicine, awarded jointly to Mario Capecchi, Martin Evans, and Oliver Smithies.

Mice are vital for basic research, but by the late 1980s, it was apparent that many practical applications in areas such as biomedicine and animal agriculture required ES technology or a functional equivalent in other species, particularly livestock. Despite considerable efforts, fully functional ES cells had not (and still have not) been derived from livestock. Despite the difficulties, transfer of nuclei from "ordinary" somatic cells was seen as a possible alternative means of transferring engineered genetic modifications from cells to whole animals. However, this would not be possible if nuclear transfer was restricted to cells obtained directly from early embryos. Genetic modification and powerful techniques like gene targeting require cells to be cultured for several weeks.

The breakthrough in nuclear transfer came in the 1990s, when it became clear that the cell-cycle stage of the nuclear donor cell must be correctly matched to the state of the recipient cytoplasm. Keith Campbell and Ian Wilmut of the Roslin Institute near Edinburgh made the key insight that only nuclei in G1 phase (prior to DNA replication) would support normal development in unfertilized oocytes. Transfer of nuclei at other stages in the cell cycle leads to aberrant chromosomal replication and inviable embryos. The method they developed and which is still widely used is to block cycling in the donor cells before DNA replication by reducing the amount of serum in the culture medium.

In 1995, two live lambs, Megan and Morag, were produced by transfer of nuclei from cells that were derived from a day 9 sheep embryo and then grown in culture for 6–13 passages [13]. The following year a similar experiment was performed that also included cultured adult sheep mammary cells. On 5 July 1996, a lamb was born from these cells and named Dolly [14]. Her birth showed that previous difficulties with nuclear transfer had indeed been technical. Adult cell nuclei were not irreversibly determined in their developmental potential and could be radically reprogrammed to adopt a new fate. Shortly after, transgenic and gene-targeted animals were produced, demonstrating that nuclear transfer from somatic cells was a practical means of generating genetically modified animals [15, 16].

Nuclear Transfer

Choice of Transgenic Methods

As mentioned in the introduction, the development of nuclear transfer in livestock mammals was largely inspired by the power of ES cell-mediated transgenesis in mice. The aim was to gain more precision and control over the type of genetic modifications that could be carried out. This section reviews nuclear transfer within the overall context of transgenic technology.

Methods of producing transgenic mammals can broadly be divided into two categories:

1. Direct transgenesis – the transfer of transgene DNA directly into embryos. Methods include DNA microinjection, viral transduction, transposon-mediated transgenesis, and sperm-mediated DNA transfer.

2. Cell-mediated transgenesis – the engineering of genetic alterations in cultured cells that are then converted to whole animals. Nuclear transfer falls within this category. The other major method is the incorporation of ES, or induced pluripotent stem cells, into developing embryos to produce chimeric animals.

Methods of direct transgenesis are more straightforward than cell-mediated transgenesis. However, at present these allow only transgene addition. Earlier techniques, such as DNA microinjection, are also very inefficient. Although not a significant drawback in mice, this is an important factor in livestock because gestation and maturation times are longer and maintenance costs far higher. For example, cattle have a generation interval between 2 and 3 years and each cow bears either one or two calves, while the generation time of mice is 3 months and litter size is between 6 and 12 pups. There has therefore been a strong incentive in livestock programs to reduce the number of animals gestating non-transgenic fetuses. Transgenic fetuses can be identified in utero by analysis of cells shed by the developing fetus and obtained by amniocentesis or allantocentesis. However, these procedures carry a significant risk of inducing abortion. Some efforts have also been made to detect and analyze fetal cells or DNA in the maternal circulation, but with little success. Definitive analysis of the integrated transgene must therefore be carried out in animals shortly after birth using small samples taken from blood, tail, or ear tips.

Methods of cell-mediated transgenesis share the important feature that genetic manipulation and analysis of the transgenic genotype are carried out in cells in the laboratory, rather than in animals “on the farm.” These cells are then used to transfer the modified genotype to whole animals. While cell-mediated transgenesis is more labor intensive than direct transgenesis, it offers significant advantages. Cells can be analyzed in detail so that only those with the desired genotype are converted to whole animals. Far larger numbers of independent transgene integration or other events can be screened and investigated in cells than in whole animals. This allows the selection and isolation of cells carrying rare integration events resulting from homologous recombination, the basis of gene targeting.

Nuclear transfer is currently the only practical form of cell-mediated transgenesis in mammals other than mice.

Nuclear Transfer Basics

The standard nuclear transfer procedure is to prepare a recipient cytoplasm by removing the genomic DNA from an unfertilized oocyte. This is usually carried out by microsurgical withdrawal of a portion of cytoplasm containing the second metaphase plate using a micromanipulator. A low-tech option is to split the oocyte under a microscope using a thin blade, dye is then used to identify the half containing the metaphase chromosomes, a procedure termed “hand made cloning.” The donor nucleus, or a whole cell, is then introduced into the enucleated oocyte by microinjection or electrofusion. Reconstructed embryos are activated with an electrical pulse to simulate fertilization, cultured if possible to identify viable embryos, and then transferred to the oviduct of foster mothers to complete gestation. Readers interested in detailed protocols are referred to a book by Verma and Trounson [17].

As mentioned in the introduction, the key insight that led to the success of nuclear transfer from a wide variety of cells was recognizing the importance of matching the cell-cycle stage of the donor nucleus and the oocyte cytoplasm. Oocytes of most mammalian species pause twice during the meiotic cell divisions that form the gametes: once before the first meiotic metaphase, and again at the second metaphase, at which stage the oocyte is mature and can be fertilized. Oocyte maturation and arrest are induced by a high level of a protein complex termed maturation-promoting factor (MPF). Fertilization causes a chain of events that result in proteolytic cleavage of MPF, breaking the arrested state and allowing the fertilized oocyte to complete meiotic division.

The level of MPF in the oocyte has a profound effect on the outcome of nuclear transfer. If a nucleus is transferred into oocyte cytoplasm with high MPF, the nuclear envelope breaks down and chromatin undergoes chromosome condensation, followed by nuclear reformation and DNA replication. A nucleus from a cell in G1 phase of the cell cycle will undergo normal DNA replication and can support normal development. However, a donor nucleus in S, or G2,

phase undergoes aberrant re-replication of DNA, causing chromosomal damage or an abnormal chromosome number with consequent failure of development. Nuclear transfer into unfertilized oocytes can therefore be improved by synchronizing donor nuclei in G1 phase. This can be achieved by depriving donor cells of serum, blocking the cell cycle before DNA replication.

It has been proposed that the key to successful reprogramming of a transferred nucleus is the free availability in the recipient cytoplasm of factors necessary for regulating the program of gene transcription [18]. These are normally associated with DNA within the nucleus, but dissociate from the genome during cell division, and are released into the cytoplasm when the nucleus breaks down, ready to be distributed with the chromosomes into the new daughter nuclei. These factors reassociate with the DNA in G1 phase to reestablish the transcriptional program. Unfertilized oocytes have an abundance of such free factors, being paused part way through meiotic metaphase. Consistent with this model, if the genome of a zygote is removed during cell division, the resulting cytoplasm can successfully reprogram an incoming cell nucleus to embryonic gene expression, but if the zygote nucleus is removed during interphase, it cannot [18].

Nuclear transfer from somatic cells opened a new route to produce transgenic livestock, and lifted the requirement for ES cells for cell-mediated transgenesis. Cells such as primary fetal fibroblasts can be obtained in large quantities, manipulated in culture, and then converted into whole animals by nuclear transfer. The initial work carried out in Edinburgh was inspired by the possibility of producing large numbers of genetically modified animals without conventional breeding, termed “instant flocks,” for applications such as biopharmaceutical production in milk [19]. The first large animal produced in this way was a Poll Dorset sheep “Polly” that carried a human clotting factor IX transgene, randomly introduced into the genome by in vitro transfection of fetal fibroblasts [15]. Shortly after, an α 1-antitrypsin transgene was placed by gene targeting into a site chosen as likely to favor expression [16]. Somatic cell nuclear transfer has since been used to generate transgenic and gene-targeted animals in several other species.

From the outset, comparative data indicated that nuclear transfer requires fewer experimental animals than DNA microinjection to produce a useful transgenic animal [15], reducing costs and benefiting animal welfare. Techniques continue to improve gradually; for example, bovine nuclear transfer data from 1998 to 1999 showed an efficiency of 6.3% calves born per reconstructed embryo, while in 2003–2004 this increased to 15% [20].

Fetal and perinatal mortality and morbidity are the most serious issues that face nuclear transfer. The severity of the problem varies between species, cell types, and experimental regimes and is unrelated to genetic manipulation of the cultured cells. Rather, it is believed to be a consequence of defective epigenetic reprogramming of the donor nucleus and possibly incompatibility between the cell-derived nuclear genome and the oocyte-derived mitochondria. Cumulative data for cattle collected up to 2005 indicate that more than 1,500 cloned calves were born, of these, 60–70% survived normally to adulthood. The performance of these, including reproduction, was similar to non-cloned animals. Evidence is also accumulating that ill effects are limited to the first generation. Offspring, including those from two nuclear transfer parents, exhibit no increased morbidity or mortality.

Nuclear Transfer with Differentiated Cells

Transgenesis by nuclear transfer began with the use of abundantly available somatic cells such as fetal fibroblasts, mainly as a matter of convenience [15]. It is desirable to characterize as fully as possible cell clones used to generate whole animals. However, most primary cell types have a short lifespan in culture, allowing little time for transfection, selection, and the expansion of individual cell clones. Furthermore, prolonged growth in culture risks genetic and epigenetic alterations that may reduce their ability to support normal development.

A number of precautions can be taken to minimize the total time cells spend in culture. Genetic manipulation should ideally be carried out using primary cells at very early passage. Samples of each cell clone can be cryopreserved and retained as a frozen stock for nuclear transfer, while a replicate sample is expanded for further analysis. Determination of the modal

chromosome number of cell clones is also useful because it allows those that have undergone gross changes in chromosome complement to be excluded. Frozen samples of cell clones identified as most suitable as nuclear donors can be thawed and used to generate whole animals. The viability and lifespan of primary cells can vary considerably between different isolates and are affected by factors such as the age of the donor animal and the method of preparation. It is therefore important to identify a suitable cell isolate before commencing a transgenic project.

The problem of short lifespan can also be overcome by “rejuvenating” cells by nuclear transfer and re-derivation from resulting fetuses. This allows successive rounds of in vitro genetic manipulation to be carried out relatively quickly, but does introduce a possible risk that genetic aberrations occur in the cultured cells, but remain undetected until animals are born. Serial nuclear transfer has been used to carry out multiple genetic manipulations in large animals. Three successive rounds of cloning were used to inactivate both alleles of the $\alpha 1,3$ -GT gene in pigs [21]. In cattle, serial nuclear transfer has been used to inactivate immunoglobulin genes and the prion protein gene PrP [22, 23]. Some researchers have described a progressive decrease in the efficiency of development over successive clonal generations (e.g., [24]). However, there has been a recent report of 15 successive generations of re-cloned mice produced with no decrease in success rate using a chromatin-modifying reagent trichostatin A [25], suggesting that animal cloning can in principle be repeated indefinitely.

Nuclear Transfer and Gene Targeting

As mentioned in Section “[History and Introduction](#),” a major motivation for the development of nuclear transfer in mammals was to circumvent the need for embryonic stem cells and enable gene targeting in species other than mice [16]. In brief, gene targeting exploits the ability of a cell to support recombination between an exogenous DNA molecule and chromosomal DNA at regions of shared homology, as part of the DNA repair process. It can be used to inactivate individual endogenous genes by insertion or deletion, replace whole genes, precisely place transgenes in the host genome, introduce subtle gene modifications, and carry out large-scale modifications such as the deletion

of megabase-sized DNA fragments. There are many variations and technical refinements, and the reader is referred to a review by Capecchi [26] for further details. Typically, cells are transfected with a DNA construct carrying an engineered modification flanked by DNA “arms” homologous to the target locus. At a certain frequency, homologous recombination occurs between the construct and the target gene and the engineered modification is seamlessly incorporated. The frequency of homologous recombination events varies by locus, vector, and the cell type used. Identifying cell clones carrying a targeted event among a large background of random integrants is key to success. Various methods have been devised to aid their isolation, the most powerful of which is to use a drug resistance marker gene that is expressed only if the vector recombines correctly at the chosen target site. Variations of this technique are referred to as “promoter trap,” “intron trap,” or “polyA trap” selection.

Nuclear transfer is currently the only practical method of producing gene-targeted livestock; however, it has become clear that gene targeting in somatic cells is considerably more difficult than targeting in mouse ES cells, and relatively few gene-targeted large animals have been generated. Genes so far targeted in livestock are: *COL1A1* in sheep [16], *PRNP* in cattle, sheep and goats [22, 27, 28], *GGTA1* in pigs [29, 30], *IGH* in cattle [22], and *CFTR* in pigs [31]. The main problems are the short lifespan of primary somatic cells in culture, as discussed above; the overall lower frequency of homologous recombination relative to random integration events in somatic cells; and the difficulty of targeting genes not expressed in the host cells.

Although direct comparisons are hard to find, it is generally understood that homologous recombination events are less frequent in somatic cells than in mouse ES cells [32]. Methods such as promoter trapping can enrich rare cell clones bearing targeted events, but lack of transcriptional activity in the host cell precludes their use. There is also evidence that spontaneous homologous recombination events are particularly rare for non-expressed genes [33]. This has made targeting of many potentially important genes difficult. There are however promising strategies to enhance the proportion of homologous recombination events in somatic cells, such as the synchronization of cells in late S phase so that vector transfection can be optimally

timed [34], but these have yet to be fully evaluated. Current progress and a review of the field are provided by Laible and Alonso-González [35].

An entirely new approach to mammalian gene targeting is now emerging, based on the use of highly specific nuclease enzymes. This promises to overcome many of these problems and possibly supplant nuclear transfer for some applications, as discussed in Section “[Beyond Nuclear Transfer – Direct Manipulation of Early Embryos](#).”

Multipotent and Pluripotent Stem Cells as Nuclear Donors

The main advantage of multipotent cells, such as mesenchymal stem cells (MSCs), and pluripotent cells, such as ES cells, over more differentiated cell types is their longer lifespan and greater stability in culture. This facilitates genetic manipulation procedures and the isolation of cell clones for the generation of transgenic animals.

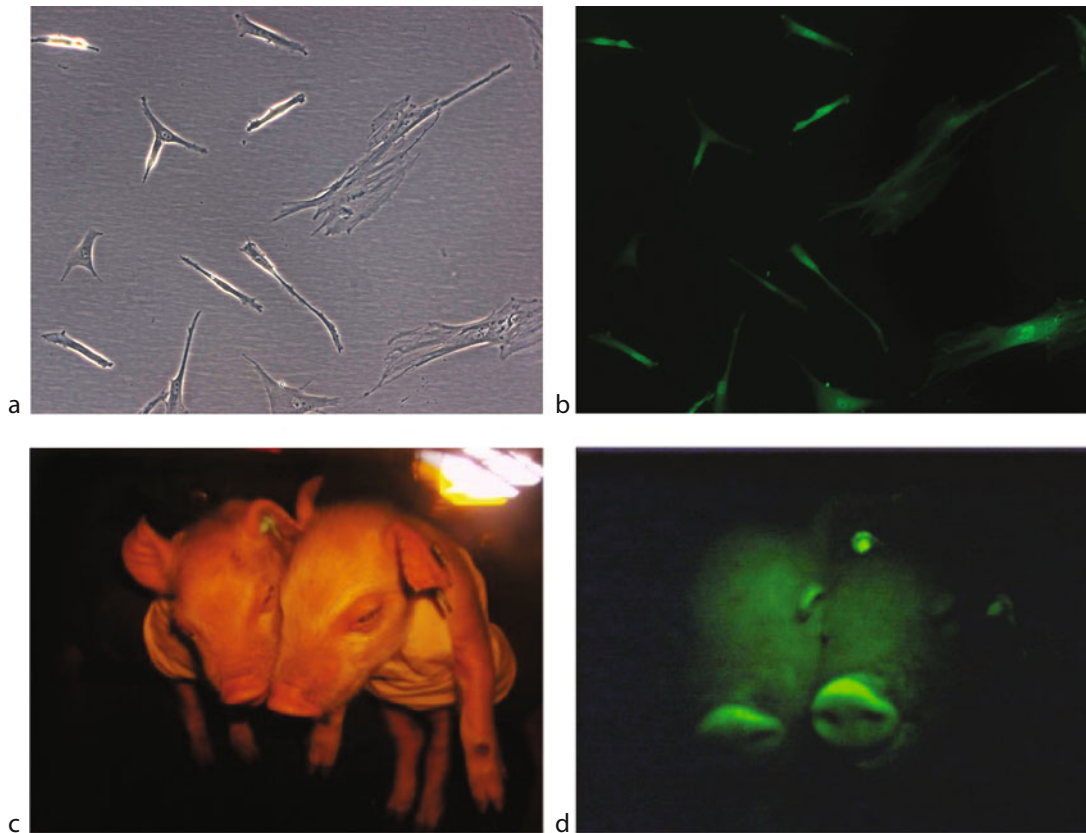
One of the defining characteristics of ES cells is that they can grow indefinitely as undifferentiated cells in culture while retaining the ability to form all tissues of the body. However, experience gained in mice is not always applicable to other species. From the few species investigated, it is clear that not all ES cells are as easy to handle as those from mouse. For example, human ES cells survive electroporation and dissociation to single cells poorly and this has made derivation of stable transfected human ES cell clones difficult [36]. Despite considerable efforts over more than a decade, few gene-targeted human ES lines have been reported [37].

However, the main problem is that ES cells are simply not available for most mammals. Artificial induction of pluripotency by the expression of key transcription factors [38, 39] offers an alternative to classical ES cells that is rapidly being extended to other species. Induced pluripotent stem (iPS) cells can be cultured for long periods, and mouse iPS cells have already been used for nuclear transfer [40]. At the time of writing, the detailed characteristics of iPS cells from most species however remain unknown. It is perhaps cautionary to note that although candidate pig iPS cells have been reported by several authors, in each case, the reprogrammed pluripotent state was not stable in culture and relied on continued expression of exogenous transcription factors.

Multipotent mesenchymal stem cells (MSCs) are attractive alternatives to ES cells for generating transgenic animals. MSCs can be readily derived from bone marrow or adipose tissue. They proliferate well for long periods in culture and have been used successfully for nuclear transfer in cattle and pigs [41–44]. MSCs can be transfected and retain their multipotent identity and support preimplantation development of nuclear transfer embryos at rates similar to non-manipulated MSCs [42, 43]. Figure 1 shows an illustrative transgenic experiment in which a green fluorescent protein reporter gene was introduced into pig MSCs and whole animals derived by nuclear transfer.

Studies of reconstructed embryos produced by nuclear transfer indicate that an important cause of developmental failure is incomplete reprogramming of the transferred nucleus. This leads to deficient

expression of early embryonic genes and failure to establish a normal embryonic pattern of epigenetic regulation and chromatin structure. It has been suggested that the epigenetic state of the donor nucleus influences reprogramming efficiency, and cells with a high degree of developmental plasticity, such as multipotent and pluripotent stem cells should be more successful nuclear donors than differentiated cell types (e.g., [45]). In mice, a higher fraction of embryos derived from ES cells were found to develop to term than those from differentiated donor cells [46]. Evidence from other species is however less clear, although some studies in pigs describe higher rates of development to blastocyst with bone marrow MSCs than fetal fibroblasts [47, 48]. These findings were supported by analysis of gene expression, which revealed that porcine MSC-derived cloned embryos



Nuclear Transfer to Produce Transgenic Mammals. Figure 1

Legend (a) gfp-transfected porcine MSCs; (b) gfp-transfected porcine MSCs under UV light; (c) nuclear transfer piglets generated from gfp MSCs; (d) nuclear transfer piglets under UV light

resembled normal embryos more closely than did those from fetal fibroblasts [48].

As mentioned in the previous section, another advantage of stem cells is that they may support homologous recombination more efficiently than other cell types, enabling gene targeting. There is insufficient data regarding MSCs, but evidence from mice indicates that iPS cells support homologous recombination at rates similar to ES cells [49]. Human ES cells also seem to support homologous recombination at a rate similar to mouse ES cells [50].

Beyond Nuclear Transfer: Direct Manipulation of Early Embryos

Sequence-specific DNA endonuclease enzymes that recognize long target sequences are new tools that promise to revolutionize genetic manipulation in many species. Enzymes of this type include naturally occurring homing endonucleases, often termed meganucleases, and artificially engineered enzymes such as zinc finger nucleases (ZFN), please see review by Cathomen and Joung [51]. Recent years have seen particular interest in the use of ZFNs in mammals.

ZFNs are artificial enzymes produced by fusing the cleavage domain of the FokI restriction endonuclease with a series of DNA-binding domains known as zinc finger motifs. Each ZFN is a heterodimer that can be produced to recognize a specific DNA target sequence of 18–24 base pairs. An appropriately designed ZFN can create a double-strand break at a single predetermined site in the genomic DNA. In eukaryotes, the error-prone non-homologous end-joining DNA repair pathway is often mutagenic, creating small insertions and deletions at the break site. Thus, careful choice of break site provides a means of inactivating a chosen gene, an alternative to gene knockout by classical gene targeting. ZFNs must be designed and custom made for each individual target site. A critical goal of enzyme design is to ensure that it does not produce double-strand breaks at sites other than that intended. Such off-target activity is likely to be toxic or mutagenic.

Perhaps the most important feature of this technology is that the frequency of gene inactivation can be so high that selection or enrichment of cells carrying the targeted event is unnecessary, effectively removing the need for cell culture. In 2008, ZFN-mediated gene inactivation was

achieved directly in vertebrate embryos. Viable gene-targeted zebra fish were created by injecting mRNAs encoding ZFNs directly into one-cell embryos [52, 53]. This approach has now been extended to mammals, with the report of rats carrying inactivated IgM and Rab38 genes produced by ZFN targeting in embryos [54].

At present, ZFNs are expensive and have yet to be widely tested, but on present evidence, it seems likely that this approach could replace somatic cell nuclear transfer as a means of generating knock out animals in many mammal species. The exception being mice, where ES cells already provide a simple means of producing gene-targeted animals. ZFNs also promise to facilitate gene-targeted insertion or substitution of defined sequences in cultured somatic cells. As mentioned in Section “[Nuclear Transfer and Gene Targeting](#),” the low frequency of homologous recombination is a problem, and some genes have been difficult to target. Evidence indicates that a double-strand break at the target site dramatically increases the frequency of targeting events [55]. It has recently been shown that ZFNs facilitate gene targeting by homologous recombination in human ES and iPS cells, which had proved difficult [56]. It remains to be seen whether gene-targeted insertions or substitutions can also be achieved directly in embryos, this is a fascinating possibility that will undoubtedly be explored in the near future.

Future Directions

Science can be very unpredictable, and history is littered with authoritative, but inaccurate forecasts of future progress. So these comments are necessarily tentative.

What is already clear is the attention that nuclear transfer has brought to questions of cell identity and fate. The year 1997 marked a radical change in experimental biology; the concept of cell determination was overturned and replaced by the possibility of complete reprogramming of cell identity. This inspired many talented people to embark on projects that were previously inconceivable, with perhaps the most notable outcome being the derivation of induced pluripotent stem cells. The production of primitive stem cells essentially at will, coupled with increasing knowledge of the normal development of many cell types, is fueling intense research activity in regenerative medicine and will undoubtedly bring significant advances in medicine and benefit to human health.

The efficiency of nuclear transfer has improved, but advances have been incremental rather than dramatic. There is still considerable mortality during gestation and around birth. First generation nuclear transfer animals sometimes face health difficulties, probably as a consequence of epigenetic defects. These may be resolved or mitigated by the use of pluripotent stem cells as nuclear donors, but supporting evidence has so far been restricted to mice. At least for a while, somatic cell nuclear transfer remains the preferred option for the production of genetically modified livestock, and the next decade will likely see gene-targeted animals of many kinds being generated, especially for biomedicine. However, while nuclear transfer remains technically difficult and inefficient, alternative methods will continually be sought. It is possible that efficient methods of embryo incorporation such as laser-assisted injection [57] could supersede nuclear transfer, with iPS cells used to generate 100% chimeras.

Another notable new technology is the use of zinc finger nucleases in early embryos. It already seems likely that this approach might take over from nuclear transfer for the production of gene knockout animals. There is also the promise that the use of ZFNs in combination with gene replacement vectors in cultured cells will significantly improve the efficiency of conventional gene targeting and give access to more difficult genes.

The availability of an enabling technology is in itself insufficient reason to produce genetically modified animals. Any transgenic project requires an objective risk to benefit analysis in which the well-being of the animal is given a high priority. But, where there are real long-term benefits to the environment, or for animal or human health, then nuclear transfer and its successor technologies can offer powerful and precise tools.

Bibliography

- Beetschen JC, Fischer JL (2004) Yves Delage (1854–1920) as a forerunner of modern nuclear transfer experiments. *Int J Dev Biol* 48:607–612
- Spemann H (1928) Die Entwicklung seitlicher und dorso-ventraler Keimhälfen bei verzögerter Kernversorgung. *Zeitschr für Wiss Zool* 132:105–134
- McGrath J, Solter D (1984) Inability of mouse blastomere nuclei transferred to enucleated zygotes to support development in vitro. *Science* 226:1317–1319
- Gilbert SF (2006) *Developmental biology*, 8th edn. Sinauer, USA, pp 573–579
- Willadsen SM (1986) Nuclear transplantation in sheep embryos. *Nature* 320:63–65
- Gordon JW, Scangos GA, Plotkin DJ, Barbosa JA, Ruddle FH (1980) Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc Natl Acad Sci USA* 77:7380–7384
- Hammer RE, Pursel VG, Rexroad CE Jr, Wall RJ, Bolt DJ, Ebert KM, Palmiter RD, Brinster RL (1985) Production of transgenic rabbits, sheep and pigs by microinjection. *Nature* 315:680–683
- Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotent cells from mouse embryos. *Nature* 292:154–156
- Martin GR (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci USA* 78:7634–7638
- Bradley A, Evans M, Kaufman MH, Robertson E (1984) Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* 309:255–256
- Smithies O, Gregg RG, Boggs SS, Koralewski MA, Kucherlapati RS (1985) Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* 317:230–234
- Thomas KR, Capecchi MR (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 51:503–512
- Campbell KH, McWhir J, Ritchie WA, Wilmut I (1996) Sheep cloned by nuclear transfer from a cultured cell line. *Nature* 380:64–66
- Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KH (1997) Viable offspring derived from fetal and adult mammalian cells. *Nature* 385:810–813
- Schnieke AE, Kind AJ, Ritchie WA, Mycock K, Scott AR, Ritchie M, Wilmut I, Colman A, Campbell KHS (1997) Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts. *Science* 278:2130–2133
- McCreath KJ, Howcroft J, Campbell KHS, Colman A, Schnieke AE, Kind AJ (2000) Production of gene-targeted sheep by nuclear transfer from cultured somatic cells. *Nature* 405:1066–1069
- Verma PJ, Trounson A (2006) *Nuclear transfer protocols: cell reprogramming and transgenesis*. Humana Press, New Jersey
- Egli D, Birkhoff G, Eggan K (2008) Mediators of reprogramming: transcription factors and transitions through mitosis. *Nat Rev Mol Cell Biol* 9:505–516
- Kind A, Schnieke A (2008) Animal pharming, two decades on. *Transgenic Res* 17:1025–1033
- Heyman Y (2005) Nuclear transfer: a new tool for reproductive biotechnology in cattle. *Reprod Nutr Dev* 45:353–361
- Phelps CJ, Koike C, Vaught TD, Boone J, Wells KD, Chen SH, Ball S, Specht SM, Polejaeva IA, Monahan JA, Jobst PM, Sharma SB, Lamborn AE, Garst AS, Moore M, Demetris AJ, Rudert WA, Bottino R, Bertera S, Trucco M, Starzl TE, Dai Y, Ayares DL (2003) Production of alpha 1, 3-galactosyltransferase-deficient pigs. *Science* 299:411–414

22. Kuroiwa Y, Kasinathan P, Matsushita H, Sathiyaseelan J, Sullivan EJ, Kakitani M, Tomizuka K, Ishida I, Robl JM (2004) Sequential targeting of the genes encoding immunoglobulin- μ and prion protein in cattle. *Nat Genet* 36:775–780
23. Kuroiwa Y, Kasinathan P, Sathiyaseelan T, Jiao JA, Matsushita H, Sathiyaseelan J, Wu H, Mellquist J, Hammitt M, Koster J, Kamoda S, Tachibana K, Ishida I, Robl JM (2009) Antigen-specific human polyclonal antibodies from hyperimmunized cattle. *Nat Biotechnol* 27:173–181
24. Kubota C, Tian C, Yang X (2004) Serial bull cloning by somatic cell nuclear transfer. *Nat Biotechnol* 22:693–694
25. Thuan NV, Kishigami S, Wakayama T (2010) How to improve the success rate of mouse cloning technology. *J Reprod Dev* 56:20–30
26. Capecchi MR (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* 6:507–512
27. Denning C, Burl S, Ainslie A, Bracken J, Dinnyes A, Fletcher J, King T, Ritchie M, Ritchie WA, Rollo M, de Sousa P, Travers A, Wilmut I, Clark AJ (2001) Deletion of the α (1, 3)galactosyl transferase (GGTA1) gene and the prion protein (PrP) gene in sheep. *Nat Biotechnol* 19:559–562
28. Yu G, Chen J, Yu H, Liu S, Chen J, Xu X, Sha H, Zhang X, Wu G, Xu S, Cheng G (2006) Functional disruption of the prion protein gene in cloned goats. *J Gen Virol* 87:1019–1027
29. Dai Y, Vaught TD, Boone J, Chen SH, Phelps CJ, Ball S, Monahan JA, Jobst PM, McCreath KJ, Lamborn AE, Cowell-Lucero JL, Wells KD, Colman A, Polejaeva IA, Ayares DL (2002) Targeted disruption of the α 1, 3-galactosyltransferase gene in cloned pigs. *Nat Biotechnol* 20:251–255
30. Lai L, Kolber-Simonds D, Park KW, Cheong HT, Greenstein JL, Im GS, Samuel M, Bonk A, Rieke A, Day BN, Murphy CN, Carter DB, Hawley RJ, Prather RS (2002) Production of α -1, 3-galactosyltransferase knockout pigs by nuclear transfer cloning. *Science* 295:1089–1092
31. Rogers CS, Hao Y, Rokhlina T, Samuel M, Stoltz DA, Li Y, Petroff E, Vermeer DW, Kabel AC, Yan Z, Spate L, Wax D, Murphy CN, Rieke A, Whitworth K, Linville ML, Korte SW, Engelhardt JF, Welsh MJ, Prather RS (2008) Production of CFTR-null and CFTR-DeltaF508 heterozygous pigs by adeno-associated virus-mediated gene targeting and somatic cell nuclear transfer. *J Clin Invest* 118:1571–1577
32. Arbonés ML, Austin HA, Capon DJ, Greenburg G (1994) Gene targeting in normal somatic cells: inactivation of the interferon- γ receptor in myoblasts. *Nat Genet* 6:90–97
33. Thyagarajan B, Johnson BL, Campbell C (1995) The effect of target site transcription on gene targeting in human cells in vitro. *Nucleic Acids Res* 23:2784–2790
34. Zaunbrecher GM, Dunne PW, Mir B, Breen M, Piedrahita JA (2008) Enhancement of extra chromosomal recombination in somatic cells by affecting the ratio of homologous recombination (HR) to non-homologous end joining (NHEJ). *Anim Biotechnol* 19:6–21
35. Laible G, Alonso-González L (2009) Gene targeting from laboratory to livestock: current status and emerging concepts. *Biotechnol J* 4:1278–1292
36. Costa M, Dottori M, Sourris K, Jamshidi P, Hatzistavrou T, Davis R, Azzola L, Jackson S, Lim SM, Pera M, Elefanti AG, Stanley EG (2007) A method for genetic modification of human embryonic stem cells using electroporation. *Nat Protoc* 2:792–796
37. Elefanti AG, Stanley EG (2009) Reshaping pluripotent stem cells. *Nat Biotechnol* 27:823–824
38. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126:663–676
39. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131:861–872
40. Kou Z, Kang L, Yuan Y, Tao Y, Zhang Y, Wu T, He J, Wang J, Liu Z, Gao S (2010) Mice cloned from induced pluripotent stem cells (iPSC). *Biol Reprod* (Epub ahead of print)
41. Kato Y, Imabayashi H, Mori T, Tani T, Taniguchi M, Higashi M, Matsumoto M, Umezawa A, Tsunoda Y (2004) Nuclear transfer of adult bone marrow mesenchymal stem cells: developmental totipotency of tissue-specific stem cells from an adult mammal. *Biol Reprod* 70:415–418
42. Bosch P, Pratt SL, Stice SL (2006) Isolation, characterization, gene modification, and nuclear reprogramming of porcine mesenchymal stem cells. *Biol Reprod* 74:46–57
43. Colleoni S, Donofrio G, Lagutina I, Duchi R, Galli C, Lazzari G (2005) Establishment, differentiation, electroporation, viral transduction, and nuclear transfer of bovine and porcine mesenchymal stem cells. *Cloning Stem Cells* 7:154–166
44. Faast R, Harrison SJ, Beebe LF, McIlpatrick SM, Ashman RJ, Nottle MB (2006) Use of adult mesenchymal stem cells isolated from bone marrow and blood for somatic cell nuclear transfer in pigs. *Clon Stem Cell* 8:166–173
45. Oback B, Wells D (2007) Donor cell differentiation, reprogramming, and cloning efficiency: elusive or illusive correlation? *Mol Reprod Dev* 74:646–654
46. Rideout WM, Eggan K, Jaenisch R (2001) Nuclear cloning and epigenetic reprogramming of the genome. *Science* 293:1093–1098
47. Kumar BM, Jin HF, Kim JG, Ock SA, Hong Y, Balasubramanian S, Choe SY, Rho GJ (2007) Differential gene expression patterns in porcine nuclear transfer embryos reconstructed with fetal fibroblasts and mesenchymal stem cells. *Dev Dyn* 236:435–446
48. Jin HF, Kumar BM, Kim JG, Song HJ, Jeong YJ, Cho SK, Balasubramanian S, Choe SY, Rho GJ (2007) Enhanced development of porcine embryos cloned from bone marrow mesenchymal stem cells. *Int J Dev Biol* 51:85–90
49. Hanna J, Wernig M, Markoulaki S, Sun CW, Meissner A, Cassady JP, Beard C, Brambrink T, Wu LC, Townes TM, Jaenisch R (2007) Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* 318:1920–1923
50. Cathomen T, Joung JK (2008) Zinc-finger nucleases: the next generation emerges. *Mol Ther* 16:1200–1207
51. Zwaka TP, Thomson JA (2003) Homologous recombination in human embryonic stem cells. *Nat Biotechnol* 21:319–321

52. Doyon Y, McCammon JM, Miller JC, Faraji F, Ngo C, Katibah GE, Amora R, Hocking TD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Amacher SL (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol* 26:702–708
53. Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* 26: 695–701
54. Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, Choi VM, Jenkins SS, Wood A, Cui X, Meng X, Vincent A, Lam S, Michalkiewicz M, Schilling R, Foeckler J, Kalloway S, Weiler H, Ménoret S, Anegón I, Davis GD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jacob HJ, Buelow R (2009) Knockout rats via embryo microinjection of Zinc-finger nucleases. *Science* 325:433
55. Smih F, Rouet P, Romanienko PJ, Jasin M (1995) Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res* 25:5012–5019
56. Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKolver RC, Katibah GE, Amora R, Boydston EA, Zeitler B, Meng X, Miller JC, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* 27:851–857
57. Poueymirou WT, Auerbach W, Frendewey D, Hickey JF, Escaravage JM, Esau L, Doré AT, Stevens S, Adams NC, Dominguez MG, Gale NW, Yancopoulos GD, DeChiara TM, Valenzuela DM (2007) F0 generation mice fully derived from gene-targeted embryonic stem cells allowing immediate phenotypic analyses. *Nat Biotechnol* 25:91–99